



**FUNDAÇÃO
UNIVERSIDADE
FEDERAL DE
MATO GROSSO DO SUL**

**FACULDADE DE ENGENHARIAS,
ARQUITETURA E URBANISMO E
GEOGRAFIA**

Engenharia Elétrica

Uso de Ferramenta de Extração e Transformação de Dados para Automação na Conciliação Fiscal

Edmilson Andrade Junior

Campo Grande MS
8 de dezembro de 2020



**FUNDAÇÃO
UNIVERSIDADE
FEDERAL DE
MATO GROSSO DO SUL**

**FACULDADE DE ENGENHARIAS,
ARQUITETURA E URBANISMO E
GEOGRAFIA**

Engenharia Elétrica

Uso de Ferramenta de Extração e Transformação de Dados para Automatização na Conciliação Fiscal

Edmilson Andrade Junior

Orientador: Prof. Dr. Cristiano Quevedo Andrea

Trabalho de Conclusão de Curso
apresentado à Universidade Federal de
Mato Grosso do Sul na Faculdade de
Engenharias, Arquitetura e Urbanismo e
Geografia, como requisito parcial para
obtenção do título de Engenheiro
Eletricista.

Campo Grande MS
8 de dezembro de 2020

Uso de Ferramenta de Extração e Transformação de Dados para Automatização na Conciliação Fiscal

Monografia apresentada à Universidade Federal de Mato Grosso do Sul na Faculdade de
Engenharias, Arquitetura e Urbanismo e Geografia, para obtenção da Graduação em
Engenharia Elétrica.

Banca Examinadora:

Prof. Dr. Cristiano Quevedo Andrea
Orientador

Prof. Me. Francisco Bayardo M. H. Barbosa

Prof. Dr. Luigi Galotto Junior

Campo Grande MS
8 de dezembro de 2020

DECLARAÇÃO DE AUTORIA E RESPONSABILIDADE

Edmilson Andrade Junior, residente e domiciliado na cidade de Campo Grande, Estado do Mato Grosso do Sul, portador do RG de nº 45.948.090 SSP/SP e CPF nº 446.660.518-13, declaro que o “Trabalho de Conclusão de Curso” apresentado, com o título “Uso de Ferramenta de Extração e Transformação de Dados para Automatização na Conciliação Fiscal” é de minha autoria e assumo a total responsabilidade pelo seu conteúdo e pela originalidade do texto. Declaro que identifiquei e referenciei todas as fontes e informações gerais que foram utilizadas para construção do presente texto. Declaro também que este artigo não foi publicado, em parte, na íntegra ou conteúdo similar em outros meios de comunicação, tendo sido enviado com exclusividade para a Universidade Federal de Mato Grosso do Sul (UFMS).

Campo Grande, 8 de dezembro de 2020.

Edmilson Andrade Junior

Dedicatória

*“Muitos são os planos no coração do homem,
mas o que prevalece é o propósito do Senhor.”*

– Provérbios 19:21

*Dedico esse trabalho especialmente aos meus
pais que sempre foram minhas referências e
sempre batalharam para que eu obtivesse as
melhores condições de estudo.*

AGRADECIMENTOS

Essa formação só se tornou possível pelas permissões grandiosas de Deus e também graças aos meus familiares e amigos próximos que sempre me apoiaram nas escolhas inflexíveis até o presente momento. Presto também, sincero agradecimento ao professor Cristiano Quevedo Andrea pela forma como sempre se predispôs a me ajudar de forma prática e objetiva, ao passo que manifesto minha gratidão a professora Flávia Zechineli por todo direcionamento e auxílio nos momentos mais críticos da graduação.

RESUMO

O departamento de contabilidade dentro de uma grande companhia é o local onde são realizadas as análises financeiras de uma dada empresa. Um de seus setores é a contabilidade fiscal, destinada a verificar todos os insumos fiscais da instituição, afim de apurar e cumprir as obrigações fiscais, evitando assim multas e punições advindas de um mau saneamento dos encargos tributários. Entretanto, há um certo gargalo comum em empresas de grande porte, situado nas verificações dos documentos fiscais, por ter alta volumetria e não haver muito esforço direcionado a essas áreas da corporação, preferindo alocar em departamentos mais estratégicos do negócio.

Diante disso, este trabalho foi pautado em um projeto realizado em uma grande multinacional, a qual proporcionou os insumos necessários para que pudesse ser feito um estudo prévio dos ganhos alcançáveis na implantação de uma automação para realizar algumas operações no departamento fiscal.

A solução implantada baseou-se em *jobs* desenvolvidos para outras finalidades dentro da companhia, mas que poderiam ser transformados para que auxiliasse na apuração destes dados fiscais. Utilizando ferramentas de extração, tratamento e cruzamento de dados, foi desenvolvida uma automação a qual importa as bases a serem comparadas, trata as informações para que fiquem preparadas para a verificação e comparação dos dados, exportando, por fim, um relatório detalhando o que foi feito com cada documento fiscal.

Os resultados obtidos foram bem satisfatórios dado os ganhos mensurados e constatados ao final do projeto realizado na companhia, mantendo a integridade dos dados e trazendo inúmeros benefícios para a área de negócio.

Palavras-Chave: Automação, companhia, contabilidade fiscal, dados, empresa, *software*.

ABSTRACT

The accounting department within a large company is where financial health is measured and evaluated. One of its sectors is tax accounting, designed to verify all the tax inputs of the institution, in order to ascertain and comply with tax obligations, thus avoiding fines and punishments from poor sanitation of tax burdens. However, there is a certain common concern in large companies, located in the verification of tax documents, for having high volume and there is not much effort directed at these areas of the corporation, preferring to allocate it in more strategic departments of the business.

Therefore, this work was based on a project carried out in a big multinational company, which provided the necessary supplies so that a preliminary study of the achievable gains in the implantation of an automation to carry out some operations in the tax department.

The implemented solution was based on jobs developed for other purposes within the company, but which could be transformed to assist in verifying these tax data. Using data extraction, processing and crossing tools, an automation was developed which imports the bases to be compared, treats the information so that they are prepared for the verification and comparison of the data, finally exporting a report detailing what was done with each tax document.

The results obtained were very satisfactory given the gains measured and verified at the end of the project carried out at the company, maintaining data integrity and bringing numerous benefits to the business area.

Keywords: Automation, company, tax accounting, data, institution, software.

SUMÁRIO

1. INTRODUÇÃO	10
1.1. Justificativa para o desenvolvimento do trabalho.....	12
1.2. Objetivos	14
1.2.1. Objetivo geral.....	14
1.2.2. Objetivos específicos	14
1.3. Organização do trabalho:	14
2. FERRAMENTAS DE ETL.....	16
2.1. <i>Extract</i>	17
2.2. <i>Transform</i>	18
2.3. <i>Load</i>	20
2.4. Extração, transformação e carregamento de dados na conciliação fiscal	21
3. REVISÃO TEÓRICA E MODELAGEM DO SISTEMA ESTUDADO	24
3.1. Revisão das bases utilizadas.....	24
3.1.1 Primeira base de dados.....	24
3.1.2 Segunda base de dados.....	26
3.1.3 Principais funcionalidades da ferramenta - Talend.....	27
3.2. Modelagem do Sistema Utilizado	33
3.2.1 Tratamento dos dados de <i>input</i>	36
3.2.2 Validação e Cruzamento dos Dados.....	42
4. RESULTADOS E DISCUSSÕES.....	47
4.1 Conclusão dos Resultados e Discussões.....	50
5. CONCLUSÕES GERAIS.....	51
6. REFERÊNCIAS	53
7. APÊNDICE A-MODELOS DE SIMULAÇÃO.....	55

1. INTRODUÇÃO

O departamento de Conciliação Fiscal nas empresas e indústrias tem como objetivo principal, assegurar que a escrituração dos lançamentos fiscais tenha sido registrada corretamente. A rotina da equipe consiste no processo de conferência das movimentações, levando em consideração as normas tributárias e visando a segurança da organização [2].

Dentre algumas das suas responsabilidades, está o controle de entrada e saída de notas fiscais da companhia, apuração de impostos, registro e escrituração de livros fiscais, executar baixas de inscrições perante as esferas municipal, estadual e federal, solicitação de negativas, dentre outros tributos e obrigações fiscais [1].

Dado que o Brasil é um dos países mais burocráticos do mundo - ocupando a 109ª posição no ranking *Doing Business* 2019 [3] dos países mais fáceis para se fazer negócios, dentre 190 países – e se destaca pela alta complexidade das suas legislações, o setor da Conciliação Fiscal tem papel fundamental no levantamento e conferência de todas evidências emitidas contra a companhia no que se refere a tributação fiscal.

De posse desses dados, a área efetua cálculos e apurações com objetivo final de escolher os devidos tributos que incidem sobre a sua simples operação, bem como fornece informações fiscais importantes para os administradores, gestores ou proprietários da empresa, buscando subsidiar a tomada de decisões por parte dessas pessoas [4].

Valendo-se do fato de que a legislação tributária do Brasil é de alta complexidade e tem constantes mudanças, é possível que esse fator implique na gestão financeira das grandes empresas. Para que isso não ocorra é necessário um grande esforço da área fiscal para que possa se atualizar e atuar de forma precisa, demandando mais tempo e esforço para atividades manuais destinadas a conferências e adaptações às alterações impostas pela legislação brasileira.

Na realidade vivenciada no mercado hoje, trabalhadores em funções burocráticas desperdiçam mais de 40% do seu dia em processos digitais manuais [5]. Com uma amostragem de 10 mil colaboradores administrativos em 11 países, a *OnePoll* – empresa europeia renomada no ramo de pesquisas qualitativas e quantitativas – apurou que 3,5 horas diárias dos trabalhadores brasileiros são ocupadas por tarefas manuais “odiadas” pelos mesmos.

Dos entrevistados, um pouco mais da metade afirma que o tipo de serviço manual e repetitivo atrapalha na realização do seu trabalho principal, muitas vezes prejudicado pela alta quantidade de horas diárias dedicadas a processos que poderiam ser facilmente automatizados [5]. A grande maioria dos trabalhadores, cerca de 85%, declaram acreditar que a automação poderia facilmente eliminar tarefas administrativas digitais manuais e repetitivas que não são o foco do seu trabalho. Destacam que com mais tempo liberado, ao empregar uma força de trabalho digital e automação das tarefas repetitivas é possível recuperar um quarto de tempo de trabalho anual (4,5 meses), poderiam desempenhar melhor seu trabalho principal, melhorar a produtividade em seu departamento e buscar oportunidades de aprender novas habilidades [5].

Visto a volumetria de horas trabalhadas em que os colaboradores poderiam estar agregando valor de outras formas mais eficientes para a companhia, aliado a alta carga tributária brasileira – cerca de um terço do Produto Interno Bruto (PIB) é destinado a pagar impostos, taxas e contribuições [6] – é necessário algum posicionamento das lideranças para que os colaboradores possam se concentrar em gerar valor ao passo que as apurações fiscais continuem sendo minuciosamente averiguadas. Nesse cenário de alta complexidade, o caminho para a escalabilidade no desenvolvimento de processos sem a perda de eficiência e precisão é, sem dúvidas, a automação.

É importante salientar e esclarecer que as automações não são implementadas com o intuito de eliminar o trabalho humano das grandes empresas. No caso em questão, a automação fiscal tem o objetivo de melhorar os recursos empresariais, para melhorar os resultados e obter dados mais precisos em muito menos tempo [6].

Muitos são os benefícios agregados ao processo de digitalização de processos com o uso de automações sendo alguns deles:

- Benefícios capturados em FTE's – *Full Time Equivalent* – com processos manuais e repetitivos sendo executados pelo robô, de forma mais rápida e eficiente que a interação humana;
- Mitigação de erros humanos, pois uma vez definido os padrões para as regras do *business*, o robô irá atuar de forma uniforme e padrão;

- Redução de custos pelo fato de a automação permitir execução no regime 24 por 7, sem férias, licenças, variações emocionais e instabilidade produtiva;
- *Compliance* fiscal dado o cenário em que um processo sendo executado automaticamente com regras bem definidas, traz maior segurança de informação e qualidade nos dados.

A automação fiscal da qual se fala no presente trabalho é referente a processos por ferramentas com interface automatizadas que contribuem com a otimização de processos manuais e repetitivos. Para o estudo em questão, tem-se a utilização de uma ferramenta de ETL, sigla em inglês para *Extract, Transform and Load*, tem sido muito utilizada e tem gerado melhorias nas atividades repetitivas e, principalmente, naquelas com grande volume de dados, pela sua alta performance quando exposta a uma grande quantidade de dados.

Esse tipo de automação tem contribuído para tornar processos mais estáveis e para que decisões importantes passem a ser tomadas com base em regras predefinidas, garantindo o *compliance* dos processos [7]. Essas automações podem ser configuradas para simular as mesmas ações executadas por seres humanos na utilização de sistemas. É possível coletar informações, extrair relatórios, tratar base de dados, otimizar fluxo de e-mails, entradas em banco de dados, entre outras programações que podem ser feitas.

A vantagem de promover a total digitalização dos processos da empresa, é o tempo da equipe sendo poupado, o custo baixo na implementação dos sistemas e interfaces, escalabilidade de operação e o alto retorno sobre o investimento. O uso de ferramentas ETL na operação fiscal permite ainda que a companhia tenha visibilidade e rastreabilidade dos acessos e modificações realizadas pela automação. A ação passa a ser exata e revela, a um clique, o caminho cursado e as atividades em cada processo. A equipe fiscal, em contrapartida, fica livre para ações, processos e trabalhos de alto valor agregado [7].

1.1. JUSTIFICATIVA PARA O DESENVOLVIMENTO DO TRABALHO

A área de contabilidade fiscal, em outras palavras, trata-se de uma atividade que busca assegurar o perfeito cumprimento de todas as exigências da legislação fiscal, sejam elas para pagamento de impostos ou envio de informações tributárias aos órgãos de arrecadação e fiscalização do governo [4].

É de conhecimento público que na atual situação de mercado que se vive, as empresas e indústrias estão sendo cada vez mais auditadas e fiscalizadas pelas entidades governamentais ou privadas. Dado essa visão é de suma importância uma operação moderna e otimizada, pois erros acontecem e existem casos em que um simples deslize acarreta sanções graves para um negócio, que incluem o pagamento de multas e bloqueios de certidões de regularidade de débitos [4].

Outro fator interessante a se considerar, o qual é importante como motivação para o presente trabalho, é a qualidade de vida dos colaboradores, que se torna de grandiosa importância quando é apontado como um dos grandes indicadores relacionados a performance no dia a dia de operações. Quando declarações de que, na atual modalidade de trabalho, quase metade dos trabalhadores sentem uma subutilização de suas habilidades ao realizar atividades manuais e repetitivas diariamente ou que o excesso de tais atividades os atrapalham em seus reais ofícios ao passo de interferirem na sua vida pessoal, com várias horas extras e sobrecarga de trabalho [5]. É de sábia decisão ser cauteloso e se inspirar nessa motivação quando se pensa em automatizar processos como esses, principalmente no caso em que 88% dos entrevistados dizem que se sentiriam mais felizes com essa mudança.

Afim de revolucionar a realidade das operações fiscais e tributárias, existem algumas formas de aprimorar a realização das atividades de tal departamento, sendo algumas delas a gestão tributária (que engloba um conjunto de ações com o intuito de administrar todos aspectos referentes a tributos) e a contratação de consultorias especializadas em tributação. Este trabalho propõe uma solução em outro pilar de otimização para o departamento: a automação fiscal tributária.

Neste sentido, o presente trabalho contém o desenvolvimento de uma automação que atua junto aos colaboradores da área fiscal e os auxilia na apuração das informações relacionadas a tributação e encargos fiscais de responsabilidade da empresa.

1.2. OBJETIVOS

1.2.1. Objetivo Geral

Desenvolver uma automação fiscal, por meio de ferramenta ETL, capaz de identificar inconsistências tributárias em relatórios internos de conciliação fiscal, com base em validações em massa feitas automaticamente em tempo real.

1.2.2. Objetivos Específicos

Primeiramente executa-se um estudo para conhecimento da realidade empresarial do departamento e mapeamento das operações realizadas manualmente e periodicamente pelos analistas da área, com o intuito de analisar intrinsecamente a aplicabilidade da automação.

Por conseguinte, é conhecido as bases utilizadas pelos colaboradores e compreendido como funciona o fluxo de trabalho para as aprovações relacionadas ao processo, de modo que, esboce-se uma solução que atenda automaticamente às necessidades do negócio e desenvolvido a automação adequada.

1.3. ORGANIZAÇÃO DO TRABALHO:

Além do capítulo introdutório atual, o restante do trabalho foi disposto da seguinte forma descrita abaixo.

No Capítulo 2, é apresentada uma contextualização das funcionalidades das ferramentas de ETL. Detalhando a aplicabilidade e as características de cada etapa dentro de projetos de automação.

No Capítulo 3, são apresentadas as bases utilizadas para o presente estudo, juntamente ao *template* utilizado na automatização e quais as validações efetuadas pela automação.

No Capítulo 4, são apresentados os resultados obtidos sobre a solução automática proposta.

No Capítulo 5, são apresentadas as conclusões finais e propostas de seguimento do projeto.

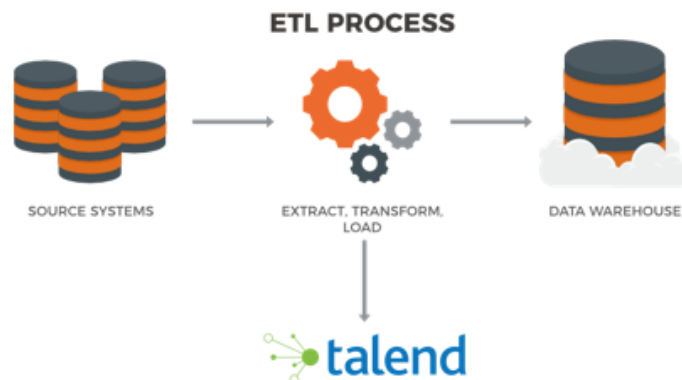
2. FERRAMENTAS DE ETL

A transformação digital é uma realidade vivenciada no mercado atual e fez com que a informação se tornasse um ativo ainda mais valioso para as organizações. Com a possibilidade de coletar e processar dados, a automação de processos atingiu um nível extremamente alto, otimizando até mesmo as tomadas de decisão dos gestores [7]. Nesse contexto, passou a ser fundamental entender o que é ETL e como tirar proveito do conceito.

As ferramentas de Extração, Transformação e Carregamento ganharam popularidade nos anos 1970, quando as organizações começaram a usar múltiplos repositórios ou bancos de dados para armazenar diferentes tipos de informações de negócios [9]. A necessidade de integrar os dados que se espalhavam pelos *databases* cresceu rapidamente. O ETL tornou-se o método padrão para coletar dados de fontes diferentes e transformá-los antes de carregar no sistema-alvo ou destino.

À medida que os *data warehouses* foram surgindo como uma nova modalidade quando se refere a banco de dados, múltiplos sistemas se tornaram capazes de integrarem acessos entre si. Com isso, *workflows*, *mainframes*, planilhas e computadores pessoais passaram a fazer parte de um único ecossistema, necessitando muitas vezes de um tratamento prévio nos dados para assim integrá-los de fato como uma grande base de dados unificada, para isso as ferramentas de ETL se tornaram extremamente úteis e eficazes e se encaixam dentro do fluxo da forma ilustrada na Figura 1.1. Para o estudo em questão a ferramenta utilizada foi o Talend [17], por ser uma ferramenta *open studio* e atende a todas exigências demandadas do projeto, importação de dados, tratamento e validação das informações por completo e geração de relatório final em diretório especificado.

Figura 1.1 Etapa na qual a ferramenta de ETL é utilizada. Fonte: Autor



Com o tempo, o número de formatos, fontes e sistemas de dados aumento em massa. Extrair, transformar e carregar é, hoje, apenas um dos vários métodos que as organizações utilizam para coletar, importar e processar dados. Há anos, inúmeras empresas têm confiado no processo de ETL para obter uma visão consolidada dos dados que geram as melhores decisões de negócios [9]. As etapas do processo são detalhadas com maiores minúcias ao longo desse capítulo.

2.1. *EXTRACT*

Esta primeira etapa do processo é destinada à coleta dos dados que farão parte do projeto [10]. Os dados são copiados ou exportados de diretórios específicos de origem até uma área de transição ou temporária. Os *databases* podem ser advindos de qualquer, estruturado ou não, base virtual – *SQL servers*, *CRMs* (*Costumer Relationship Management*), sistemas *ERPs* (*Enterprise Resource Planning*), arquivos de texto, e-mails, páginas *web*, *APIs* (*Application Programming Interface*), entre outros.

Em muitos casos, essa fase do processo é a mais crítica e desafiadora, pelo simples fato da extração correta dos dados definir o cenário para o andamento dos processos subsequentes do projeto [11]. Devido a diversidade de aplicação para esta etapa, o objetivo geral da fase de extração é converter os dados em um único formato que é apropriado para a transformação (próximo passo) em processamento.

É importante salientar uma parte intrínseca da extração de dados, que envolve a análise de dados extraídos, resultando em uma verificação se os dados atendem a um padrão ou estrutura esperada. Caso contrário, os dados podem ser rejeitos total ou parcialmente [11].

2.2. TRANSFORM

Assim que a extração de dados for concluída, já na transformação das bases, é importante uma prévia fase de tratamento e limpeza das informações. A complexidade desse bloco de consolidação dos dados que devem ser utilizados para as validações posteriores, há de depender do número de fontes na extração, do grau de heterogeneidade e dos erros nas bases [12].

No atual estágio do processo é comum haver algum tipo de correção em problemas de instância de fonte única, divergências na integração de esquemas e dados, limpeza de instâncias de múltiplas fontes, exclusão de dados duplicados, incompatibilidade de informações, tratamento de dados nulos, eliminação de linhas e/ou colunas as quais não têm valor agregado para o *business*, afim de reduzir a base ao máximo para uma otimização na performance da ferramenta. É de grande valia que essa etapa não seja negligenciada, pelo fato de preparar os *databases* para uma perfeita integração e uma conseguinte validação ao cruzar os dados junto às regras de negócio mapeadas no início do projeto.

Após a limpeza dos dados ser concluída, ainda dentro da transformação dos dados, a etapa de validação e cruzamento das bases é elaborada. É a fase onde uma série de regras ou funções são aplicadas para os dados extraídos das origens, afim de derivá-los para o carregamento no destino final. Algumas fontes de dados exigem poucas ou escassas manipulações de dados. Em outros casos, uma ou várias transformações e validações, desde básicas até avançadas, são necessárias para uma correta conciliação [11].

Podemos descrever algumas transformações como manipulações básicas caso contenha as seguintes características [12]:

- Padronização de formato: uniformizar o tipo e o comprimento dos dados de acordo com o formato do campo para facilitar a recuperação dos dados pelo usuário final;

- Ordenação: inclui o mapeamento de valores para algum significado derivado ou curto, como a transformação de ‘Masculino’ para ‘M’, ‘nulo’ para ‘0’, etc;
- Eliminação duplicatas: envolve a remoção de dados e valores duplicados nas bases;
- Implementação de restrições: estabelecimento de relacionamentos chave nas tabelas.

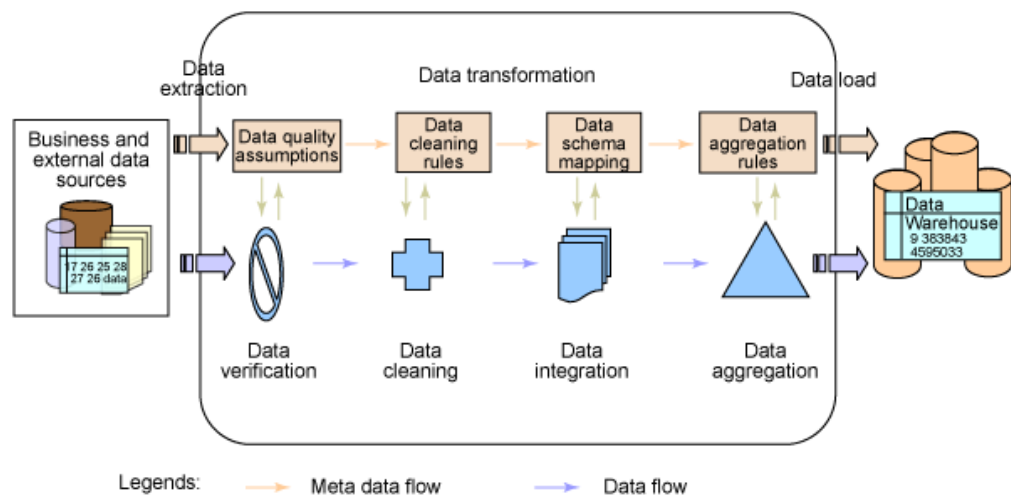
Como descrito acima, conta-se também com as transformações avançadas dentro da ferramenta de automação, as validações mais complexas seguem os seguintes padrões:

- Decodificação de campos: dados provenientes de fontes múltiplas muitas vezes serão descritos por valores de campo variáveis e, muitas vezes, sistemas de fonte legada usam códigos bastante encriptados para representar valores de negócios, tornando necessário a remoção de campos com informações semelhantes e/ou alteração de códigos obscuros em valores que fazem senso comercial para os usuários que consomem os dados;
- Mesclagem de informações: mescla de campos relacionados e visualização dos mesmos como uma única entidade, por exemplo, produto, preço do produto, tipo de produto, descrição, etc;
- Divisão de campos únicos: divisão de grandes campos de texto em campos desmembrados e únicos, por exemplo, dividindo o nome completo em *first_name*, *middle_name* e *last_name*;
- Valores calculados e derivados: em alguns casos, faz-se necessário uma agregação no conjunto de dados antes de carregá-los em um *data warehouse*, por exemplo, calcular o custo total, margem de lucro, alíquota total incidente sobre os produtos, etc;
- Resumo: os valores são resumidos para obter um valor total que é subsequentemente calculado e armazenado em vários níveis como fato de negócios em tabelas multidimensionais.

A etapa de transformação pode ser arquitetada e estruturada de diversas formas diferentes, dependendo das demandas a serem atendidas no projeto. Será visto a frente que há possibilidades de fasear a operação desse estágio da automação, orquestra-lo para operar em paralelo ou em cascata, incluir dependências relacionadas a conclusão correta (ou incorreta) de componentes ou até sub-jobs (que são exemplos de “mini projetos” dentro dos projetos maiores).

A exímia importância da etapa em detalhamento pode ser ilustrada na Figura 2.1, onde observa-se o exato local onde ela se encaixa.

Figura 2.1 Fase *transform* dentro do processo de execução das ETLs. Fonte: [13]



2.3. LOAD

Como última etapa do processo da ETL, a fase de *load* faz o carregamento dos dados no destino final, que pode ser desde um diretório específico de acesso da área de negócios (ex: pasta no *desktop*) até um banco de dados estruturado (*Data Warehouse*).

Dependendo dos requisitos da organização, o carregamento dos dados varia amplamente. Alguns *Data Warehouses* podem substituir as informações existentes com informações cumulativas, com atualizações frequentes de dados de extração, diariamente, semanalmente ou mensalmente [11]. O tempo e o escopo para substituir e anexar dados são escolhas de *design* estratégico que dependem do tempo disponível e as necessidades do negócio. Sistemas mais complexos podem manter um histórico e trilha de auditoria de todas as mudanças nos dados carregados no *Data Warehouse*.

O carregamento de dados é uma etapa muito importante do processo e pode conter várias aplicabilidades intrínsecas ao fato de simplesmente carregar as informações em algum diretório definido, como [11]:

- Por exemplo, uma instituição financeira pode ter informações sobre um cliente em vários departamentos e cada departamento pode ter as informações desse cliente listadas de uma maneira diferente. O departamento de filiação listar o cliente por nome, enquanto o departamento da contabilidade pode listar o cliente por número. Na ferramenta ETL pode-se agrupar todos esses dados e consolidá-los em uma apresentação uniforme, como para um armazenamento em banco de dados;
- Outra maneira que as grandes empresas usam a ferramenta é para mover informações para outra aplicação permanente. Por exemplo, um novo aplicativo que será utilizado usa um fornecedor de banco de dados diferente do atual, ou seja, a estrutura e *script* padrão provavelmente são distintos do atual. A ETL pode ser usada para transformar os dados em um formato adequado de carregamento para a nova aplicação do projeto;
- Outro exemplo de geração de relatórios a partir do carregamento dos dados seria um Sistema de Recuperação de Despesas e Custos, como usado pelas contabilidades, consultorias e advogados. Os dados geralmente acabam no sistema de prestação de contas, embora algumas empresas também utilizem os dados brutos deste tipo de relatório para acompanhar a produtividade de seus colaboradores com os Recursos Humanos.

2.4. EXTRAÇÃO, TRANSFORMAÇÃO E CARREGAMENTO DE DADOS NA CONCILIAÇÃO FISCAL

Alguns dos benefícios de automatizar as operações fiscais das grandes empresas foram citados nesse trabalho na sessão introdutória de número 1, porém tem-se grandes

vantagens especificamente no uso das ferramentas de ETL implementadas em conjunto às execuções diárias da área de conciliação fiscal.

Dado que o presente trabalho foi inspirado em um projeto executado em uma das quatro maiores *tradings* de *commodities* do mundo, serão explicitadas algumas das principais melhorias obtidas e também o que é possível fazer ao implementar o uso de ferramenta ETL para auxiliar nas atividades manuais e repetitivas da rotina da área de negócio.

No projeto realizado na modalidade *Summer Job*, foi elaborado o estudo de *business case*, esboço da solução, desenvolvimento da automação, homologação e validação da solução dentro de dois meses – período de férias de verão. A priori o entendimento e mapeamento do processo como um todo foi de grande importância pelo fato da mensuração dos ganhos serem estimados de acordo com a cadeia de valor do cliente.

Quando se fez a análise das validações realizadas pelos analistas na rotina de operação diária foi esboçado uma possível solução a qual atenderia às necessidades e também agregaria o máximo de valor possível para a área de negócios. Uma simulação do projeto realizado será especificada no capítulo seguinte, mas algumas das apurações efetuadas pelas automações ao comparar dois relatórios (um interno e outro externo) fornecidos, são:

- Números das NF-e (nota fiscal eletrônica);
- Números das séries das notas fiscais;
- CNPJ dos fornecedores;
- CNPJ das filiais da companhia;
- Quantidades corretas dos itens das notas fiscais;
- Valores totais das notas fiscais;
- CFOPs (Código Fiscal de Operações e Prestações) dos itens contidos nas notas fiscais;
- Cobrança indevida de impostos (ICMS, PIS, COFINS e FECOEP).

Diante do escopo do projeto realizado, a automação atuava todos os dias pela manhã. Na operação comum dos analistas da área, os mesmos perdiam cerca de 3 a 4 horas diárias, período que se estendia quando o departamento entrava em *freezing* para fechamento de mês, na apuração destes dados. De forma repetitiva e manual, os colaboradores eram expostos ao erro humano, facilitado pela grande volumetria de dados a serem conferidos diariamente.

Após a implementação das automações, essa verificação era feita a partir de um clique em que o analista selecionava o arquivo *.bat* executável, que, em poucos minutos executava as conferências programadas e definidas pela área de negócio e exportava um relatório final apenas com as inconsistências encontradas ao longo do processo de validação com os devidos direcionamentos, indicando qual o problema em cada linha da execução e auxiliando já com uma tratativa recomendada para cada tipo de caso específico.

A cada tipo de projeto é necessário o estudo e entendimento do processo *end-to-end*, para entender-se o que gera mais valor para o cliente em questão. Dessa forma é possível mapear a solução já customizada o máximo possível para que, as validações e conferência dos dados sejam feitas da forma mais eficaz possível, ou seja, a cada tipo de proposta é realizável a decisão junto a área de negócios quais as validações a serem realizadas pela automação, assim podendo agregar valor ao *business*.

3. REVISÃO TEÓRICA E MODELAGEM DO SISTEMA ESTUDADO

O projeto foi desenvolvido no ambiente da ferramenta *Talend*. A seguir serão apresentadas algumas das funcionalidades básicas do *software* utilizado, passando pelos principais componentes utilizados durante o desenvolvimento da solução.

Além disso, será apresentado um modelo semelhante às bases utilizadas no presente projeto, demonstrando a estrutura dos relatórios utilizados para o cruzamento de dados. Ao passo que os *inputs* são expostos, a solução completa, passando por cada sessão de validação e apuração das informações, será exibida em detalhes ao longo desse capítulo.

Por fim, um exemplo do reporte final esperado pela automação será demonstrado detalhando as colunas validadas e o racional das apurações efetuadas.

3.1. REVISÃO DAS BASES UTILIZADAS

Esta sessão visa detalhar duas bases semelhantes às bases originais utilizadas para realização do projeto, as quais são confidenciais por conterem informações tributárias e de cunho fiscal sigiloso.

A ideia é que as bases possam ser extraídas por analistas da área de negócio e dessa forma sejam alocadas em um diretório específico e já parametrizado dentro da interface da automação. Assim, as bases são importadas no projeto de forma automática e a partir desse passo as validações seguem automaticamente até o reporte final gerado em uma outra pasta especificada e parametrizada pelo *business*.

3.1.1 PRIMEIRA BASE DE DADOS

A primeira base de dados utilizada para o projeto em questão é um exemplo de relatório extraído internamente pela companhia. O qual contém a chave de acesso das notas fiscais, que é a chave única do processo, os valores finais das notas (com e sem a incidência dos impostos), os dados CFOP's (Código Fiscal de Operações e Prestações das entradas e saídas de mercadorias intermunicipal e interestadual) de cada nota, quantidade de itens, valor

3 – Revisão Teórica e Modelagem do Sistema Estudado

especificado por item, dentre outras informações mais específicas de uso interno da companhia.

Abaixo segue uma ilustração da base citada como tabela e uma breve detalhada das colunas de dados para entendimento:

Tabela 1 Primeira base de dados utilizada no projeto Fonte: Autor

invoice_key	num_nf	serie_nf	num_item	cod_item	cod_cfop	item_amount	item_price_tax	item_price_wtax
666796586	23564	3	1	6901	5101	12	R\$ 750.768,90	R\$ 715.018,00
666796586	23564	3	2	8250	5101	10	R\$ 1.047.839,10	R\$ 997.942,00
666796586	23564	3	3	4410	5101	5	R\$ 105.222,60	R\$ 100.212,00
869118612	12587	5	1	1065	5401	18	R\$ 526.756,65	R\$ 501.673,00
661187273	45879	4	1	1029	5403	23	R\$ 297.785,25	R\$ 283.605,00
510804431	56987	1	1	6193	5102	9	R\$ 1.061.813,88	R\$ 983.161,00
510804431	56987	1	2	7000	5102	15	R\$ 438.889,32	R\$ 406.379,00
352572494	25964	2	1	4907	6101	7	R\$ 474.723,72	R\$ 439.559,00
517639823	23648	8	1	5154	5201	8	R\$ 269.457,84	R\$ 249.498,00
588982809	20548	7	2	8876	5201	32	R\$ 964.199,16	R\$ 892.777,00
171267229	98512	2	1	7226	6102	27	R\$ 1.073.188,44	R\$ 993.693,00
793449696	12546	6	1	3949	5104	21	R\$ 164.012,94	R\$ 160.797,00
882027439	65824	8	1	9350	5151	17	R\$ 847.586,34	R\$ 830.967,00
562245232	20367	9	1	2046	5152	36	R\$ 626.372,82	R\$ 614.091,00
763564033	56982	4	1	4817	5401	28	R\$ 929.053,44	R\$ 829.512,00
420954339	89520	2	1	8254	5202	13	R\$ 722.108,80	R\$ 644.740,00
420954339	89520	2	2	6048	5202	11	R\$ 655.039,84	R\$ 584.857,00
280719603	56321	8	1	6856	5402	14	R\$ 633.446,24	R\$ 565.577,00
587028655	85224	4	1	8211	5209	24	R\$ 247.570,40	R\$ 221.045,00
587028655	85224	4	2	5355	5209	32	R\$ 775.354,72	R\$ 692.281,00

Invoice_key: essa coluna contém a chave de acesso da nota fiscal, uma chave única que será usada na identificação e cruzamento das bases.

Num_nf: coluna que contém o número da nota fiscal da nota, usada para identificação, porém não é dado único.

Serie_nf: número da série das notas fiscais.

Num_item: coluna que identifica qual o número do item em sua dada nota fiscal.

Cod_item: código de identificação do item no sistema ERP da companhia.

Cod_cfop: coluna onde contém o código CFOP do item especificado.

Item_amount: quantidade do item na nota fiscal.

Item_price_tax: valor total do item com os impostos incidentes no item.

Item_price_wtax: valor total do item sem os impostos incidentes no item.

3.1.2 SEGUNDA BASE DE DADOS

A segunda base de dados utilizada no processo é um relatório extraído do site da SEFAZ (Secretaria de Estado da Fazenda) o qual exporta as informações referentes às notas fiscais emitidas contra a companhia por uma periodicidade mensal. No entanto, nesse relatório as informações não vêm detalhadas por item, a taxa de impostos e valores das notas são referentes à nota como um todo, ou seja, esta base é o consolidado dos itens de cada nota fiscal.

Neste caso, é necessário fazer um tratamento prévio na primeira base para unificar os dados e as duas bases ficarem na mesma granularidade e assim as validações fazerem sentido para o processo. Etapa que será detalhada mais à frente no capítulo 3.2.2 onde há a demonstração dos tratamentos das bases e validações das informações.

Abaixo segue uma ilustração da base extraída da SEFAZ, com dados mais concisos e gerais das notas fiscais e uma breve detalhada das colunas de dados para entendimento:

Tabela 2 Segunda base de dados utilizada no projeto Fonte: Autor

invoice_id	nf_id	serie_id	amount_id	total_value_invoice_tax	total_value_invoice
666796586	23564	3	27	R\$ 1.903.830,60	R\$ 1.813.172,00
562245232	20367	9	36	R\$ 626.372,82	R\$ 614.091,00
869118612	12587	5	18	R\$ 526.756,65	R\$ 501.673,00
661187273	45879	4	22	R\$ 297.785,25	R\$ 283.605,00
510804431	56987	1	24	R\$ 1.500.703,20	R\$ 1.389.540,00
352572494	25964	2	7	R\$ 474.723,72	R\$ 439.559,00
352572494	25964	2	7	R\$ 474.723,72	R\$ 439.559,00
517639823	23648	8	8	R\$ 269.457,84	R\$ 249.498,00
588982809	20548	7	32	R\$ 964.199,16	R\$ 892.777,00
171267229	98512	2	27	R\$ 1.073.188,44	R\$ 993.693,00
793449696	12546	6	21	R\$ 179.967,78	R\$ 176.439,00
256487961	15628	8	12	R\$ 63.720,00	R\$ 59.000,00
965832130	26996	1	24	R\$ 146.845,44	R\$ 135.968,00
965832130	26996	1	24	R\$ 146.845,44	R\$ 135.968,00
420360517	75013	6	9	R\$ 31.270,32	R\$ 28.954,00
882027439	65824	8	17	R\$ 847.586,34	R\$ 830.967,00
562245232	20367	9	36	R\$ 626.372,82	R\$ 614.091,00
763564033	56982	4	28	R\$ 929.053,44	R\$ 829.512,00
420954339	89520	2	24	R\$ 1.377.148,64	R\$ 1.229.597,00
965832130	26996	1	24	R\$ 146.845,44	R\$ 135.968,00
280719603	56321	8	14	R\$ 633.446,24	R\$ 565.577,00
587028655	85224	4	74	R\$ 1.347.648,84	R\$ 1.247.823,00

Invoice_id: chave de acesso da nota fiscal emitida pela Secretaria Estadual da Fazenda.

Nf_id: número da nota fiscal.

Serie_id: número da série da nota fiscal.

Amount_id: quantidade total de itens na nota fiscal.

Total_value_invoice_tax: valor total da nota fiscal com os impostos incidentes da nota fiscal.

Total_value_invoice: valor total da nota fiscal sem os impostos incidentes sobre a nota fiscal.

3.1.3 PRINCIPAIS FUNCIONALIDADES DA FERRAMENTA - TALEND

Como foi falado anteriormente, no capítulo 2, o *Talend Open Studio* é uma ferramenta *Open Source* (gratuita) utilizada para tratamento e cruzamento de dados. Pela sua amplitude e diversidade de funcionalidades, é um *software* muito utilizado em grandes projetos e companhias por conta da gama de aplicabilidades em que ele se encaixa nos desenvolvimentos.

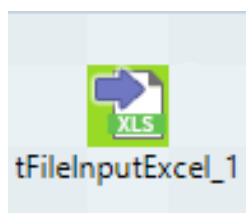
Alguns de seus principais componentes foram utilizados no presente projeto e também suas principais lógicas de validação para apuração dos dados, algumas funcionalidades essenciais estão listadas abaixo:

- Arquivos de *input*

Existem alguns componentes que são utilizados para fazer a entrada dos dados dentro da ferramenta, são alguns deles: *tFileInputExcel*, *tFileInputDelimited*, *tFileInputJSON*, *tFileInputRegex*, *tFileInputXML* e o *tFileList*.

I. *tFileInputExcel*

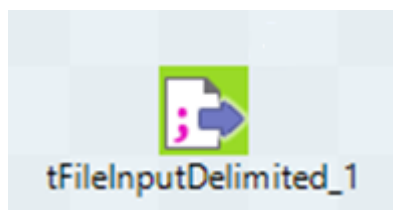
Figura 3.1 Componente de *input* relacionado à arquivo Excel Fonte: Talend



Muito usado nos projetos, é importante pelo fato de alguns relatórios, planilhas do *business*, de-para's serem importados em arquivos .xls ou .xlsx e assim o Talend consegue lê-los e tratá-los de uma forma simples e eficaz.

II. *tFileInputDelimited*

Figura 3.2 Componente de *input* relacionado à arquivos .txt ou .csv Fonte: Talend



O input de arquivos em TXT e CSV é muito importante pelo fato de várias extrações, de diversos sistemas serem feitas em um arquivo CSV ou até mesmo TXT.

Como o Talend não contém um componente exclusivamente dedicado a arquivos de input destas extensões, que é o caso do Excel com o tFileInputExcel, é utilizado uma função um pouco mais genérica, o tFileInputDelimited, no qual combinado a um *schema*, se forma um *input* da maneira desejada do usuário.

III. *tFileInputJSON*

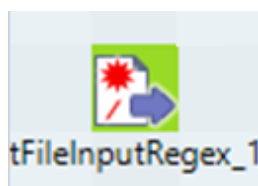
Figura 3.3 Componente de *input* relacionado à arquivos. json Fonte: Talend



Este componente tem como objetivo extrair dados JSON de um arquivo e, em seguida, transferir os dados para um arquivo, uma tabela de banco de dados etc.

IV. *tFileInputRegex*

Figura 3.4 Componente de *input* relacionado à arquivos. regex Fonte: Talend



Recurso poderoso que requer algum conhecimento avançado sobre sintaxe de expressão regular.

Abre um arquivo e o lê linha por linha para dividi-los em campos usando expressões regulares. Em seguida, envia os campos definidos no *schema* escolhido para o próximo componente do *job*.

V. *tFileInputXML*

Figura 3.5 Componente de *input* relacionado à arquivos .xml Fonte: Talend



Componente com a função de criar um arquivo estruturado XML e lê-lo linha por linha para dividi-los em campos e, em seguida, envia os campos conforme definido no *schema* para o próximo componente.

VI. *tFileList*

Figura 3.6 Componente de *input* relacionado à vários arquivos Fonte: Talend



Para lidar com vários arquivos, pode ser feita uma importação em massa, ou seja, reúne-se todos arquivos a serem *inputados* em uma pasta.

De acordo com a máscara informada, o componente itera sobre todos os arquivos de um diretório específico definido pelo usuário e carrega as linhas para o *job*.

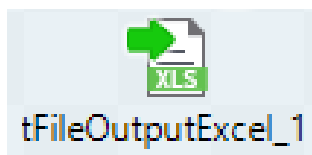
Podendo ser uma alternativa ao input em CSV e TXT.

- Arquivos de *output*

Alguns componentes dentro da ferramenta são destinados, com uma formatação definida, para a saída de dados ou carregamento dos mesmos em determinados diretórios especificados dentro do projeto, são alguns deles: *tFileOutputExcel*, *tFileOutputDelimited*, *tFileOutputJSON* e *tFileOutputXML*.

I. *tFileOutputExcel*

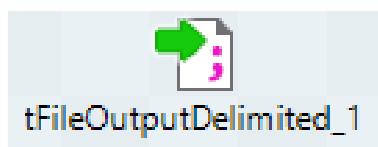
Figura 3.7 Componente de *output* relacionado à arquivo Excel Fonte: Talend



O *tFileOutputExcel* é o componente dedicado ao carregamento dos dados cruzados no *job* em uma planilha excel. Bastante utilizados nos projetos, exportar um relatório consolidado com as tratativas feitas dentro do Talend pode ser uma fácil e prática opção quando é necessário o *double check* da área de negócio.

II. *tFileOutputDelimited*

Figura 3.8 Componente de *output* relacionado à arquivos .txt e .csv Fonte: Talend

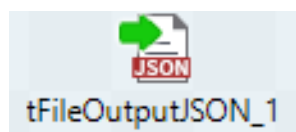


O *tFileOutputDelimited* é o componente dedicado a exportação dos dados validados dentro do *job* em um arquivo delimitado da escolha do usuário. A opção mais indicada quando é necessário um arquivo CSV ou TXT como documento final.

Combinado a um *schema* bem definido, o arquivo pode ser formatado da melhor forma possível, adequando-se às necessidades do *job* em questão.

III. *tFileOutputJSON*

Figura 3.9 Componente de *output* relacionado à arquivos .json Fonte: Talend



O *tFileOutputJSON* é usado para reescrever os dados recebidos pelo componente em um arquivo estruturado no formato JSON.

IV. *tFileOutputXML*

Figura 3.10 Componente de *output* relacionado à arquivos .xml Fonte: Talend



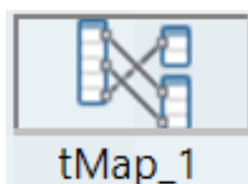
O componente *tFileOutputXML* deve ser utilizado quando pretende-se gravar um arquivo XML com valores de dados separados de acordo com o *schema* definido, que são recebidos de outros componentes anteriores.

- Componentes de desenvolvimento e validação

Existem alguns componentes com configurações pré-definidas os quais são específicos para determinadas validações, tratamento e cruzamento de dados. Serão exemplificados alguns dos que foram utilizados ao longo do presente projeto.

I. *tMap*

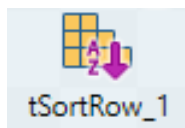
Figura 3.11 Componente de cruzamento de dados *tMap* Fonte: Talend



O *tMap* é a principal ferramenta para transformação de dados do Talend. Com ele conseguimos manipular os dados de uma forma bem fácil e intuitiva. Podemos executar correção de valores, criação de colunas, consolidação de duas fontes de dados e muito mais.

II. *tSortRow*

Figura 3.12 Componente de tratamento de dados *tSortRow* Fonte: Talend

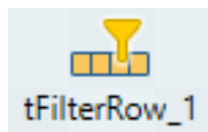


A função deste elemento é classificar os dados de entrada com base em uma ou várias colunas, por tipo de classificação e ordem. Assim auxiliando na criação de métricas e tabela de classificação.

É um componente que lida com fluxo de dados, portanto, requer entrada e saída, assim é definido como um bloco de atividades intermediário.

III. *tFilterRow*

Figura 3.13 Componente de tratamento de dados *tFilterRow* Fonte: Talend

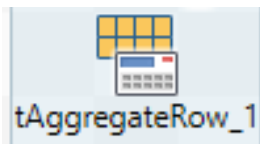


A função principal do elemento é filtrar as linhas de entrada que chegam do componente anterior e definir algumas condições customizáveis para algumas colunas selecionadas.

Precisa de um componente de input e outro de output, é possível fazer lógicas combinatórias para filtrar os dados da forma desejada.

IV. *tAggregateRow*

Figura 3.14 Componente de validação de dados *tAggregateRow* Fonte: Talend

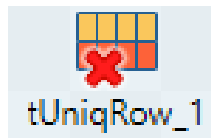


Este elemento recebe um fluxo e o agrega com base em uma ou mais colunas. Para cada linha de saída, são fornecidas a chave de agregação e o resultado relevante das operações definidas (mínimo, máximo, soma, entre outros).

É um componente que lida com fluxo de dados, portanto, requer entrada e saída, assim é definido como um bloco de atividades intermediário.

V. *tUniqueRow*

Figura 3.15 Componente de tratamento de dados *tUniqueRow* Fonte: Talend

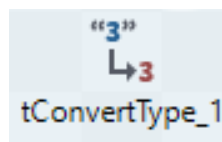


Elemento que tem como função compara entradas e classifica as entradas duplicadas do fluxo de input. Garantindo assim a qualidade dos dados de fluxo de entrada ou saída de um *job*, mitigando a duplicidade de dados.

É um componente que lida com fluxo de dados, portanto, requer entrada e saída, assim é definido como um bloco de atividades intermediário.

VI. *tConvertType*

Figura 3.16 Componente de validação de dados *tConvertType* Fonte: Talend



Elemento que permite conversões específicas em tempo de execução de um tipo java Talend para outro. Conversão de *strings*, *int*, *float*, *double*, entre outros. Este componente não pode ser usado como um componente inicial, pois requer um fluxo de entrada para operar.

3.2. MODELAGEM DO SISTEMA UTILIZADO

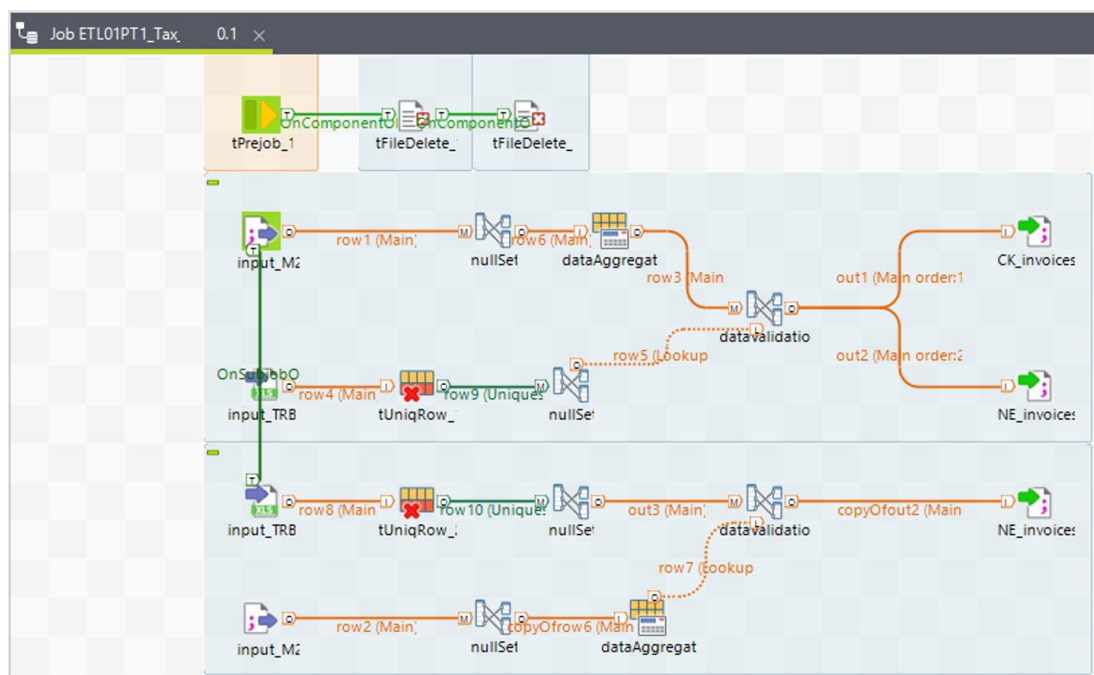
A seguinte sessão foi destinada para o detalhamento do tratamento dos dados e validações efetuadas ao longo do projeto, com a finalidade de demonstrar a importância e a aplicabilidade do presente projeto em processos os quais são utilizados mão de obra manual e repetitiva por parte de analistas e afins.

Teremos três etapas na presente sessão, sendo elas divididas em:

1. Tratamento dos dados de *input*: Onde será enfatizado quais os tratamentos feitos nas bases de dados fornecidas pela companhia em questão, para que os dados pudessem ser validados sem serem corrompidos, duplicados e até infieis;
2. Cruzamento e validações dos dados: Parte relativamente mais crítica do projeto, onde a automação faz a apuração dos dados, validando cada uma das regras definidas e impostas dentro de cada componente pertinente para os tipos de verificações que estão presentes no projeto;
3. *Outputs* esperados do projeto: Por fim, será apresentado um protótipo de saída de dados, que é o resultado das verificações e o que escolhemos, sempre alinhados com o *business*, para ser o resultado de todas as validações efetuadas ao longo do desenvolvimento.

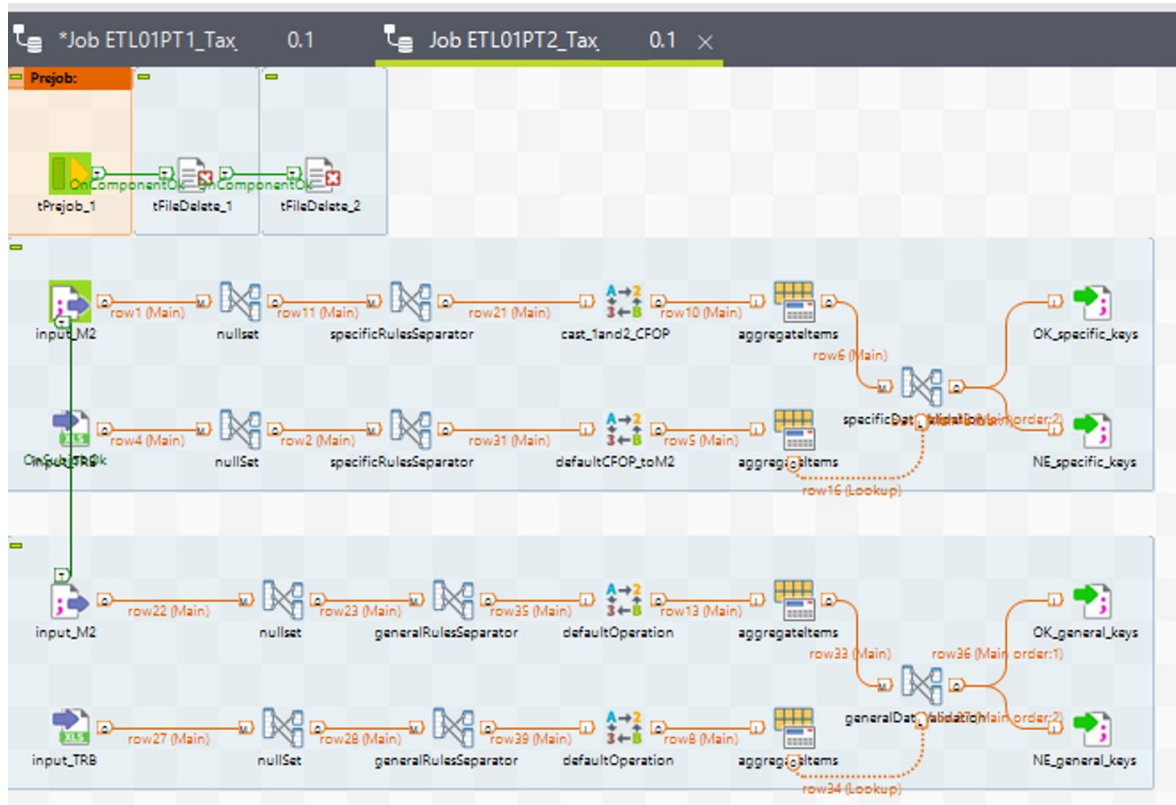
Tendo em vista que o processo completo é o conjunto dos dois *jobs* ilustrados nas Figuras 3.17 e 3.18, será detalhado abaixo cada um dos passos executados pela automação para o entendimento preciso da aplicabilidade de cada componente da ferramenta.

Figura 3.17 Primeiro *Job* efetuado pela automação Fonte: Talend



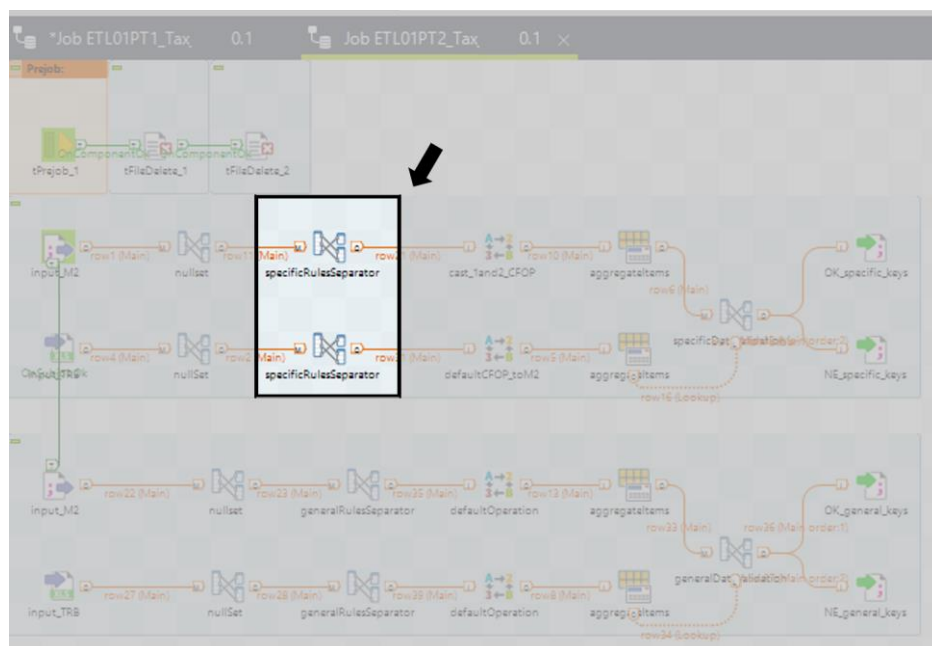
3 – Revisão Teórica e Modelagem do Sistema Estudado

Figura 3.18 Segundo *Job* efetuado pela automação Fonte: Talend



Vale ressaltar que os componentes detalhados por etapa serão ilustrados na Figura 3.19, pois como indicado abaixo, o segundo *job* trata-se de uma solução customizada desenvolvida para o cliente, o qual foi necessário algumas informações específicas e confidenciais.

Figura 3.19 Destaque das informações confidenciais/específicas do segundo *job* Fonte: Talend

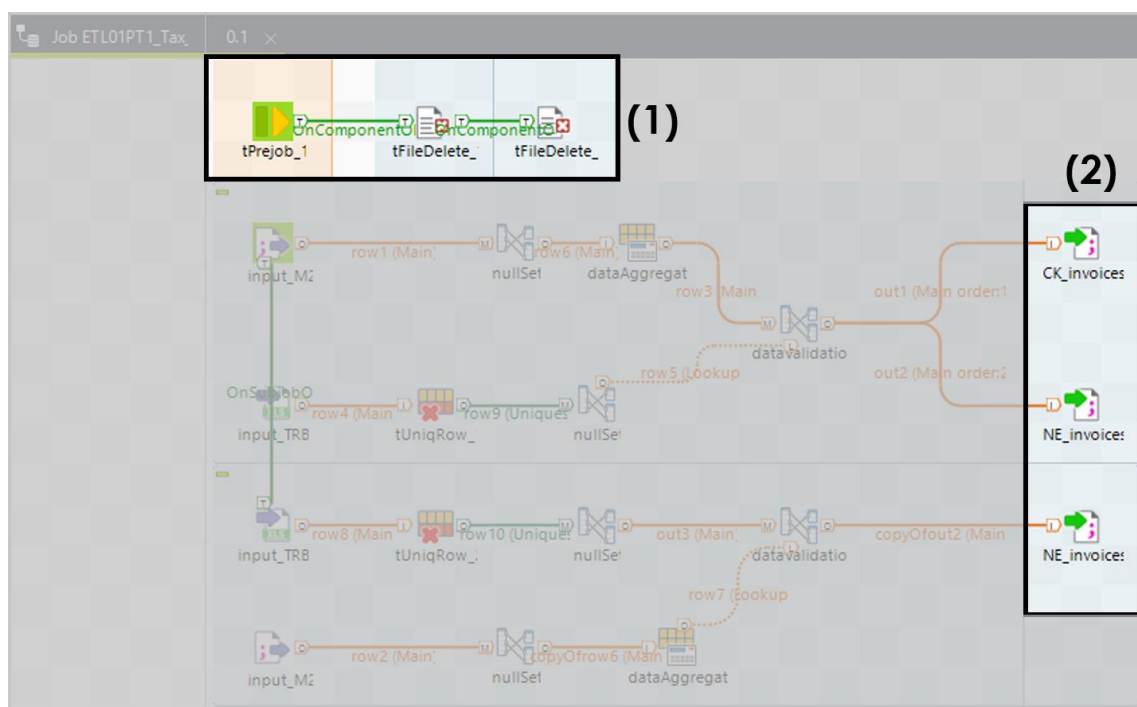


3.2.1 TRATAMENTO DOS DADOS DE INPUT

No tratamento de dados de *input* as bases foram tratadas antes de passarem para a esteira de verificação, este processo é detalhado nos itens seguintes:

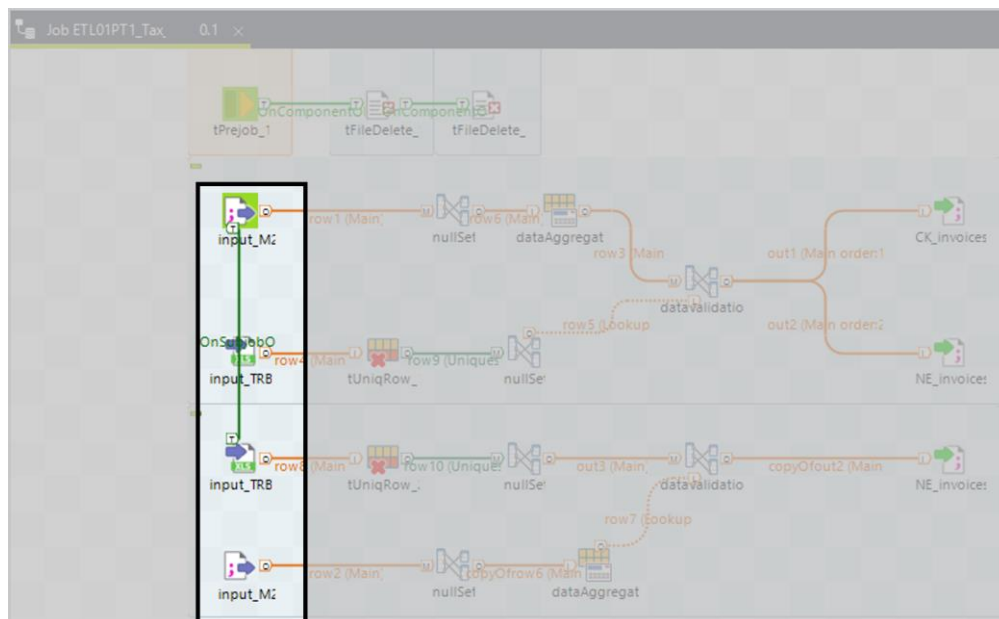
- a. No intuito de não haver duplicação de dados e como o consumo dos relatórios gera. ados é diário, temos um *pré-job* que executa antes do funcionamento do restante do *job*. Neste caso temos um componente chamado *tFileDelete* demonstrado na Figura 3.20 (1) que é responsável por verificar, no diretório passado como parâmetro, se há algum arquivo com o nome os relatórios destacados na Figura 3.20 (2). Se houver, o componente tem a missão de deletá-los antes de iniciar as validações do *job*, para que não aja duplicidade de dados.

Figura 3.20 Componente *tFileDelete* sendo executado como *pré-job* Fonte: Talend



- b. Após a verificação e remoção de arquivos existentes com mesmo nome registrados nos relatórios finais, seguimos com a importação dos arquivos iniciais, as duas bases detalhadas algumas sessões acima, as quais serão utilizadas para o tratamento de dados posteriores:

Figura 3.21 Componentes de *input* sendo importados para tratamento Fonte: Talend



- c. Em seguida às importações das bases utilizadas como mostrado na Figura 3.21, iremos tratá-las. Em um primeiro momento, a base número 2 passa pelo componente *tUniqRow* (Figura 3.22) que tem a função de eliminar linhas duplicadas, pois foi identificado na base 2 que haviam algumas notas fiscais que saíam duplicadas na extração pelo sistema da Secretaria Estadual da Fazenda (Figura 3.40). Com a missão de transformar o arquivo importado em uma base de dados confiável, o componente verifica quais linhas contém os mesmos valores em todas as colunas e as exclui, deixando somente uma das linhas válida (Tabelas 3 e 4).

Figura 3.22 Componente de exclusão de linhas duplicadas na base Fonte: Talend

3 – Revisão Teórica e Modelagem do Sistema Estudado

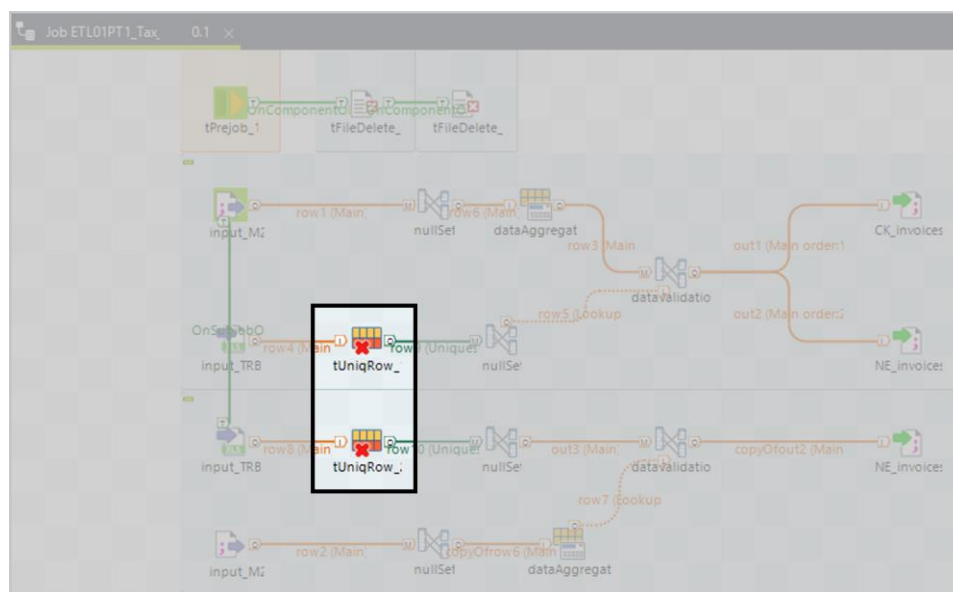


Tabela 3 Base 2 com dados duplicados e destacados para auxílio de identificação

Fonte: Autor

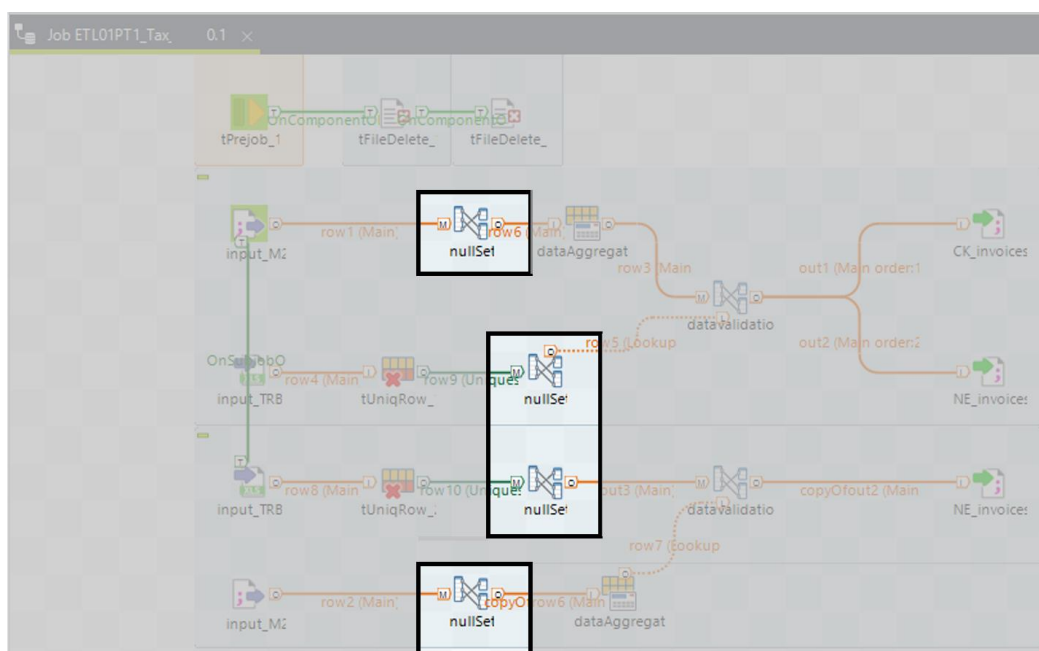
invoice_id	nf_id	serie_id	amount_id	total_value_invoice_tax	total_value_invoice
666796586	23564	3	27	R\$ 1.903.830,60	R\$ 1.813.172,00
562245232	20367	9	36	R\$ 626.372,82	R\$ 614.091,00
869118612	12587	5	18	R\$ 526.756,65	R\$ 501.673,00
661187273	45879	4	22	R\$ 297.785,25	R\$ 283.605,00
510804431	56987	1	24	R\$ 1.500.703,20	R\$ 1.389.540,00
352572494	25964	2	7	R\$ 474.723,72	R\$ 439.559,00
352572494	25964	2	7	R\$ 474.723,72	R\$ 439.559,00
517639823	23648	8	8	R\$ 269.457,84	R\$ 249.498,00
588982809	20548	7	32	R\$ 964.199,16	R\$ 892.777,00
171267229	98512	2	27	R\$ 1.073.188,44	R\$ 993.693,00
793449696	12546	6	21	R\$ 179.967,78	R\$ 176.439,00
256487961	15628	8	12	R\$ 63.720,00	R\$ 59.000,00
965832130	26996	1	24	R\$ 146.845,44	R\$ 135.968,00
965832130	26996	1	24	R\$ 146.845,44	R\$ 135.968,00
420360517	75013	6	9	R\$ 31.270,32	R\$ 28.954,00
882027439	65824	8	17	R\$ 847.586,34	R\$ 830.967,00
562245232	20367	9	36	R\$ 626.372,82	R\$ 614.091,00
763564033	56982	4	28	R\$ 929.053,44	R\$ 829.512,00
420954339	89520	2	24	R\$ 1.377.148,64	R\$ 1.229.597,00
965832130	26996	1	24	R\$ 146.845,44	R\$ 135.968,00
280719603	56321	8	14	R\$ 633.446,24	R\$ 565.577,00
587028655	85224	4	74	R\$ 1.347.648,84	R\$ 1.247.823,00

Tabela 4 Base 2 após passar pelo tratamento do componente Fonte: Autor

invoice_id	nf_id	serie_id	amount_id	total_value_invoice_tax	total_value_invoice
666796586	23564	3	27	R\$ 1.903.830,60	R\$ 1.813.172,00
869118612	12587	5	18	R\$ 526.756,65	R\$ 501.673,00
661187273	45879	4	22	R\$ 297.785,25	R\$ 283.605,00
510804431	56987	1	24	R\$ 1.500.703,20	R\$ 1.389.540,00
352572494	25964	2	7	R\$ 474.723,72	R\$ 439.559,00
517639823	23648	8	8	R\$ 269.457,84	R\$ 249.498,00
588982809	20548	7	32	R\$ 964.199,16	R\$ 892.777,00
171267229	98512	2	27	R\$ 1.073.188,44	R\$ 993.693,00
793449696	12546	6	21	R\$ 179.967,78	R\$ 176.439,00
256487961	15628	8	12	R\$ 63.720,00	R\$ 59.000,00
420360517	75013	6	9	R\$ 31.270,32	R\$ 28.954,00
882027439	65824	8	17	R\$ 847.586,34	R\$ 830.967,00
562245232	20367	9	36	R\$ 626.372,82	R\$ 614.091,00
763564033	56982	4	28	R\$ 929.053,44	R\$ 829.512,00
420954339	89520	2	24	R\$ 1.377.148,64	R\$ 1.229.597,00
965832130	26996	1	24	R\$ 146.845,44	R\$ 135.968,00
280719603	56321	8	14	R\$ 633.446,24	R\$ 565.577,00
587028655	85224	4	74	R\$ 1.347.648,84	R\$ 1.247.823,00

- d. Em seguida, os dados passam por uma verificação corriqueira e bem simples, onde é identificado se há algum valor vazio na base, se sim o componente foi desenvolvido para que preencha com *null* como mostrado na Figura 3.23.

Figura 3.23 Componente que verifica os valores vazios das bases Fonte: Talend



- e. O tratamento da próxima etapa é feito na base número 1. Por ser a base extraída do sistema ERP (*Enterprise Resource Planning*) da companhia, é uma base mais detalhada, com alguns dados a mais, além de ser na granularidade item a item de cada nota fiscal (Tabela 5). Como a base de dados da SEFAZ é em nível de nota fiscal, ou seja, cada linha da base é uma nota fiscal, necessita-se deixar a base 1 no mesmo grau de comparação para que a validação seja feita corretamente posteriormente. Este componente *tAggregateRow* (Figura 3.24) tem a função de, a depender de algumas “colunas chave” definidas pelo usuário serem iguais, é possível escolher o que é feito com as demais colunas da base. Neste caso foram utilizadas as colunas “invoice_key”, “num_nf” e “serie_nf” como colunas chave, por ser um dado único por nota fiscal e as colunas “item_amount”, “item_price_tax” e “item_price_wtax” serem somadas, e algumas colunas (“num_item”, “cod_item” e “cod_cfop”) que não eram pertinentes para a validação o momento foram excluídas pelo componente. Então quando a automação identificar que há uma chave de acesso igual, ele irá somar as quantidades e valores da nota fiscal, tornando assim a base na mesma granularidade da base 2 (Tabela 4), uma linha por documento, possibilitando uma validação certa.

Tabela 5 Base 1 detalhada a nível item e destacados para auxílio de identificação Fonte:

Autor

invoice_key	num_nf	serie_nf	num_item	cod_item	cod_cfop	item_amount	item_price_tax	item_price_wtax
666796586	23564	3	1	6901	5101	12	R\$ 750.768,90	R\$ 715.018,00
666796586	23564	3	2	8250	5101	10	R\$ 1.047.839,10	R\$ 997.942,00
666796586	23564	3	3	4410	5101	5	R\$ 105.222,60	R\$ 100.212,00
869118612	12587	5	1	1065	5401	18	R\$ 526.756,65	R\$ 501.673,00
661187273	45879	4	1	1029	5403	23	R\$ 297.785,25	R\$ 283.605,00
510804431	56987	1	1	6193	5102	9	R\$ 1.061.813,88	R\$ 983.161,00
510804431	56987	1	2	7000	5102	15	R\$ 438.889,32	R\$ 406.379,00
352572494	25964	2	1	4907	6101	7	R\$ 474.723,72	R\$ 439.559,00
517639823	23648	8	1	5154	5201	8	R\$ 269.457,84	R\$ 249.498,00
588982809	20548	7	1	8876	5201	32	R\$ 964.199,16	R\$ 892.777,00
171267229	98512	2	1	7226	6102	27	R\$ 1.073.188,44	R\$ 993.693,00
793449696	12546	6	1	3949	5104	21	R\$ 164.012,94	R\$ 160.797,00
882027439	65824	8	1	9350	5151	17	R\$ 847.586,34	R\$ 830.967,00
562245232	20367	9	1	2046	5152	36	R\$ 626.372,82	R\$ 614.091,00
763564033	56982	4	1	4817	5401	28	R\$ 929.053,44	R\$ 829.512,00
420954339	89520	2	1	8254	5202	13	R\$ 722.108,80	R\$ 644.740,00
420954339	89520	2	2	6048	5202	11	R\$ 655.039,84	R\$ 584.857,00
280719603	56321	8	1	6856	5402	14	R\$ 633.446,24	R\$ 565.577,00
587028655	85224	4	1	8211	5209	24	R\$ 247.570,40	R\$ 221.045,00
587028655	85224	4	2	5355	5209	32	R\$ 775.354,72	R\$ 692.281,00

3 – Revisão Teórica e Modelagem do Sistema Estudado

Figura 3.24 Componente que agrega os dados conforme definido Fonte: Talend

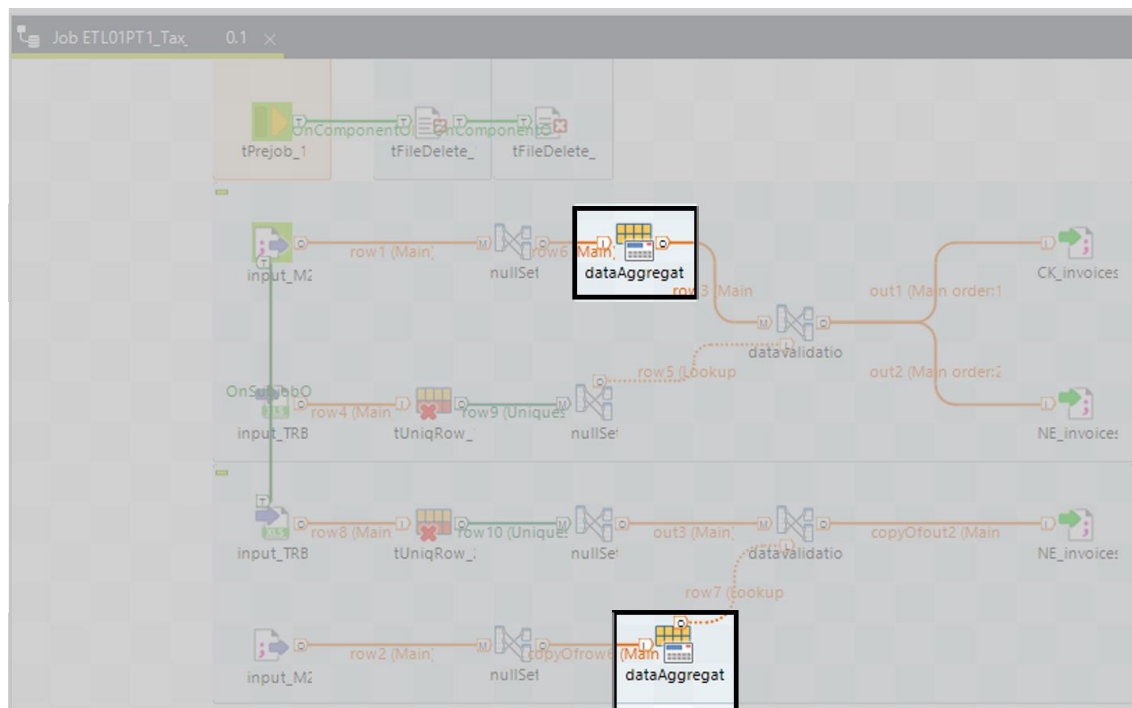


Tabela 6 Base de dados 1 após agregação por item Fonte: Autor

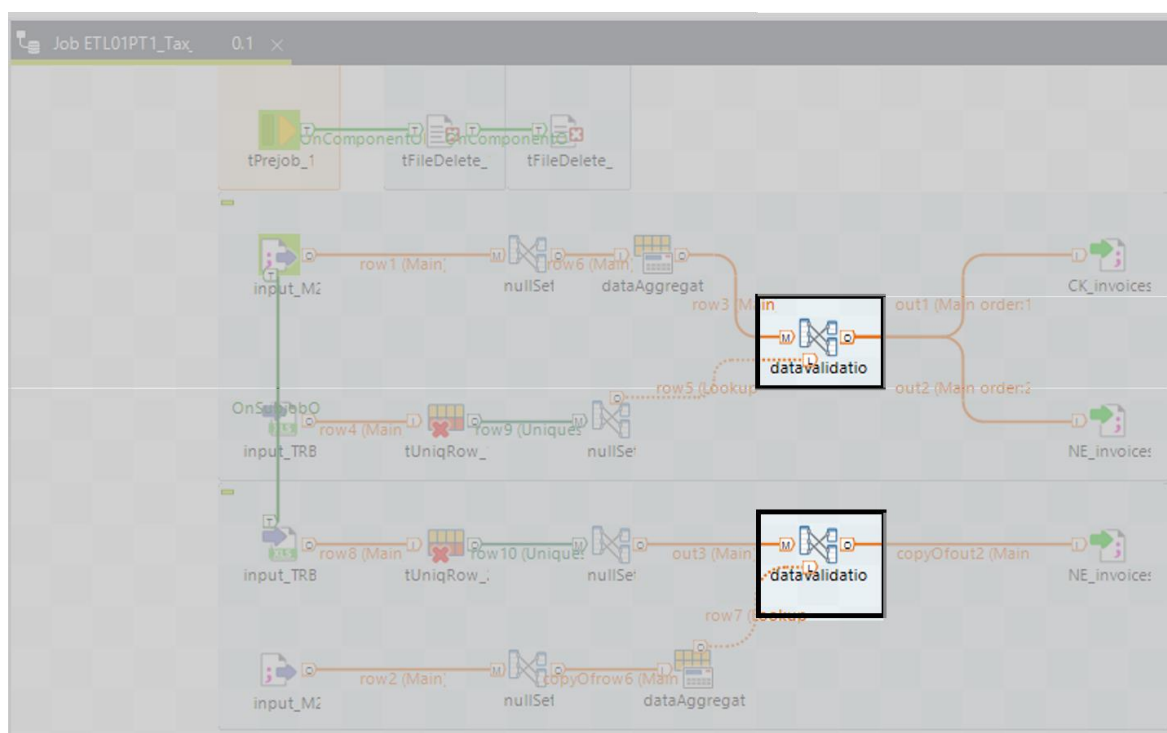
invoice_key	num_nf	serie_nf	item_amount	item_price_tax	item_price_wtax
666796586	23564	3	27	R\$ 190.383.060,00	R\$ 1.813.172,00
869118612	12587	5	18	R\$ 526.756,65	R\$ 501.673,00
661187273	45879	4	23	R\$ 297.785,25	R\$ 283.605,00
510804431	56987	1	24	R\$ 150.070.320,00	R\$ 1.389.540,00
352572494	25964	2	7	R\$ 474.723,72	R\$ 439.559,00
517639823	23648	8	8	R\$ 269.457,84	R\$ 249.498,00
588982809	20548	7	32	R\$ 964.199,16	R\$ 892.777,00
171267229	98512	2	27	R\$ 1.073.188,44	R\$ 993.693,00
793449696	12546	6	21	R\$ 164.012,94	R\$ 160.797,00
882027439	65824	8	17	R\$ 847.586,34	R\$ 830.967,00
562245232	20367	9	36	R\$ 626.372,82	R\$ 614.091,00
763564033	56982	4	28	R\$ 929.053,44	R\$ 829.512,00
420954339	89520	2	24	R\$ 137.714.864,00	R\$ 1.229.597,00
280719603	56321	8	14	R\$ 633.446,24	R\$ 565.577,00
587028655	85224	4	56	R\$ 1.022.925,12	R\$ 913.326,00

3.2.2 VALIDAÇÃO E CRUZAMENTO DOS DADOS

Este é considerado um dos momentos mais importantes e críticos da automação, onde é verificado se os relatórios das notas fiscais interno da companhia está de acordo com o relatório extraído da Secretaria Estadual da Fazenda, para que, dentro do prazo estabelecido pela SEFAZ sejam feitas as correções pertinentes, assim evitando o pagamento de multas referentes a inconsistências tributárias (estaduais e nacionais) encontradas nas notas fiscais.

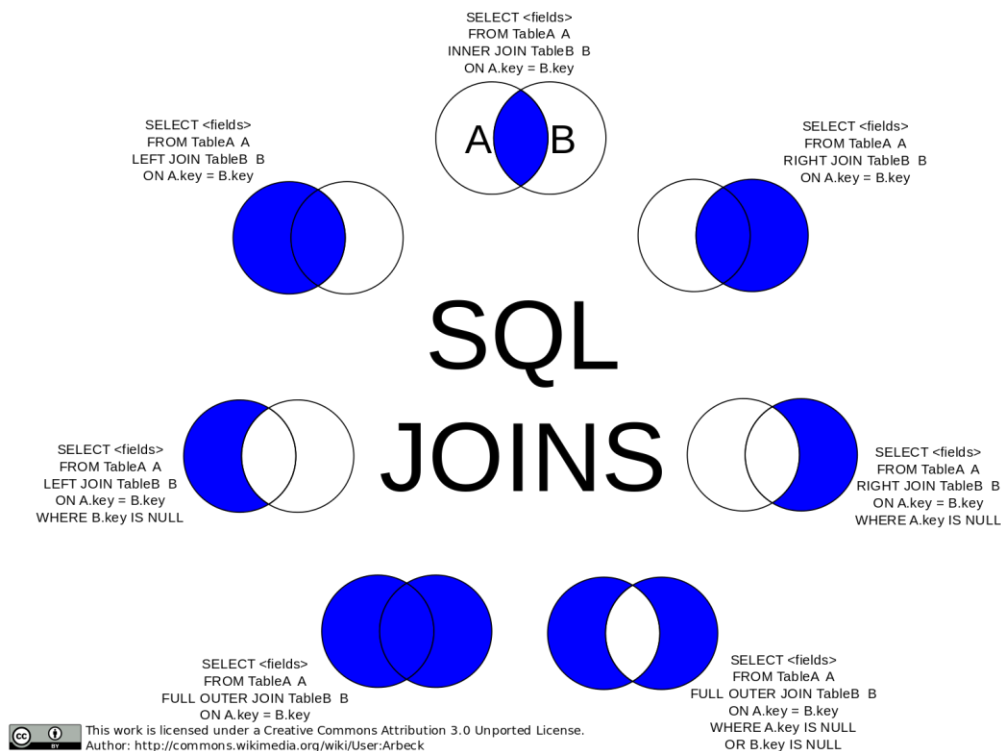
Todas as apurações são feitas no componente *tMap* indicado na Figura 3.25, no qual é utilizado dois dele, pelo fato de necessitarmos de um cruzamento *full outer join*.

Figura 3.25 Componente *tMap* onde são feitas as verificações do *job* Fonte: Talend



Dentro do componente é possível fazer algumas relações de tabela, o que se faz necessário para um cruzamento de dados eficaz. É compreendido nas linguagens de programação e banco de dados que existem alguns tipos de relação entre tabelas, elas estão exemplificadas na Figura 3.26.

Figura 3.26 Possibilidades de *joins* Fonte: [14]



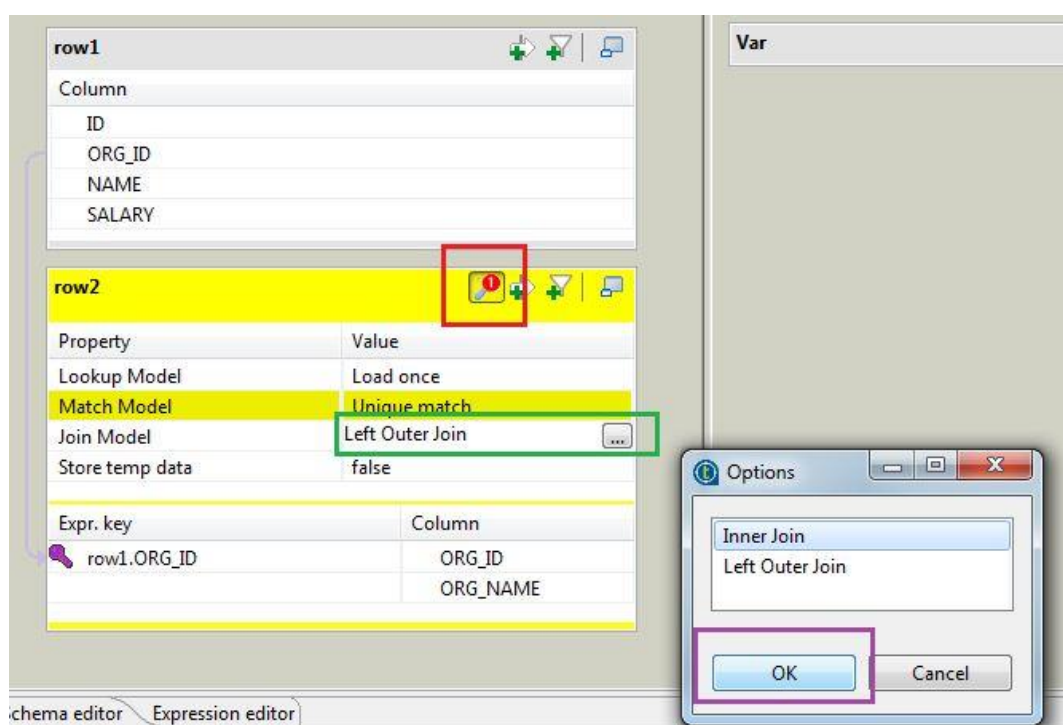
Algumas das relações entre tabelas ilustradas na Figura 3.26 são:

- *Left Join / Left Outer Join:* É quando a tabela “da esquerda” que significa a tabela primária é relacionada com uma segunda tabela e o *join* delas resulta em todos componentes que elas têm em comum mais os componentes contidos em uma dada tabela 1 que não estão na tabela 2;
- *Right Join:* Definido pela tabela “da direita” que significa a tabela secundária é relacionada com uma primeira tabela e o *join* resultante é todos os componentes que elas têm em comum mais os componentes contidos em uma dada tabela 2 que não estão presentes na tabela 1;
- *Inner Join:* Quando duas tabelas são relacionadas e o *join* delas resulta somente nos componentes existentes nas duas tabelas em comum;
- *Full Outer Join:* No relacionamento entre duas tabelas é trazido todos os dados: as informações em comum, o que contém na tabela 1, mas não existe na tabela 2 e o que há na tabela 2, porém não é encontrado na tabela 1.

Para o presente projeto, foi necessário trabalhar com o *Full Outer Join*, pois era preciso, além de validar os valores das notas encontradas em comum nos dois relatórios, identificar o que havia subido para o sistema interno, mas não para a SEFAZ e também as notas contidas no relatório da Secretaria Estadual da Fazenda, porém não eram encontradas internamente.

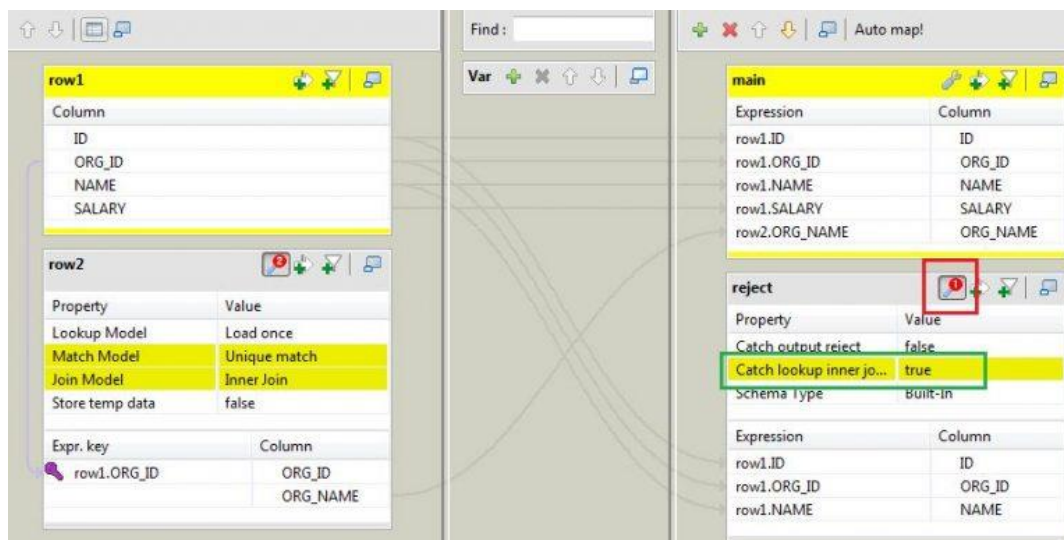
Como o componente *tMap* só possibilita as relações de *Inner Join* e *Left Outer Join* (Figura 3.27) foi necessário utilizar dois dele para completar a operação desejada de *Full Outer Join*.

Figura 3.27 Tipos de *joins* aceitos pelo componente *tMap* Fonte: [15]



Como já explicado anteriormente no início da sessão 3.2.2, o componente *tMap* contém diversas funcionalidades, inclusive trabalhar com duas saídas, ou seja, entram duas bases para verificação e pode-se trabalhar com uma saída de dados “pertinentes” às validações a serem feitas e um outro *output* com dados que não são interessantes para as validações, pode-se ver na Figura 3.28 a saída “*reject*” que seriam os dados que não são significativos para as apurações.

Figura 3.28 Pode-se trabalhar com até duas saídas no componente *tMap* Fonte: [15]



Dessa forma os dois componentes foram organizados e configurados para resultarem nesse tipo de saída de dados, de forma que se soubesse quais notas fiscais estavam em um dos sistemas, mas não estavam no outro.

O primeiro *tMap* da Figura 3.29 (1) foi utilizado da seguinte forma:

- O cruzamento das bases foi feito no componente e para as notas comumente encontradas foi exportado um relatório “CK_invoices” da Figura 3.29 (2) onde é verificado a quantidade de itens, o valor final da nota com os impostos e o valor total sem a incidência dos impostos.
- Ainda no mesmo componente, a base número 1 foi usada como base primária e usando a funcionalidade de “saídas aceitas” e “saídas rejeitadas”, exportou-se um relatório chamado “NE_invoices” da Figura 3.29 (3) a fim de obter as notas fiscais faturadas pela companhia, porém não encontradas no relatório extraído da Secretaria Estadual da Fazenda.

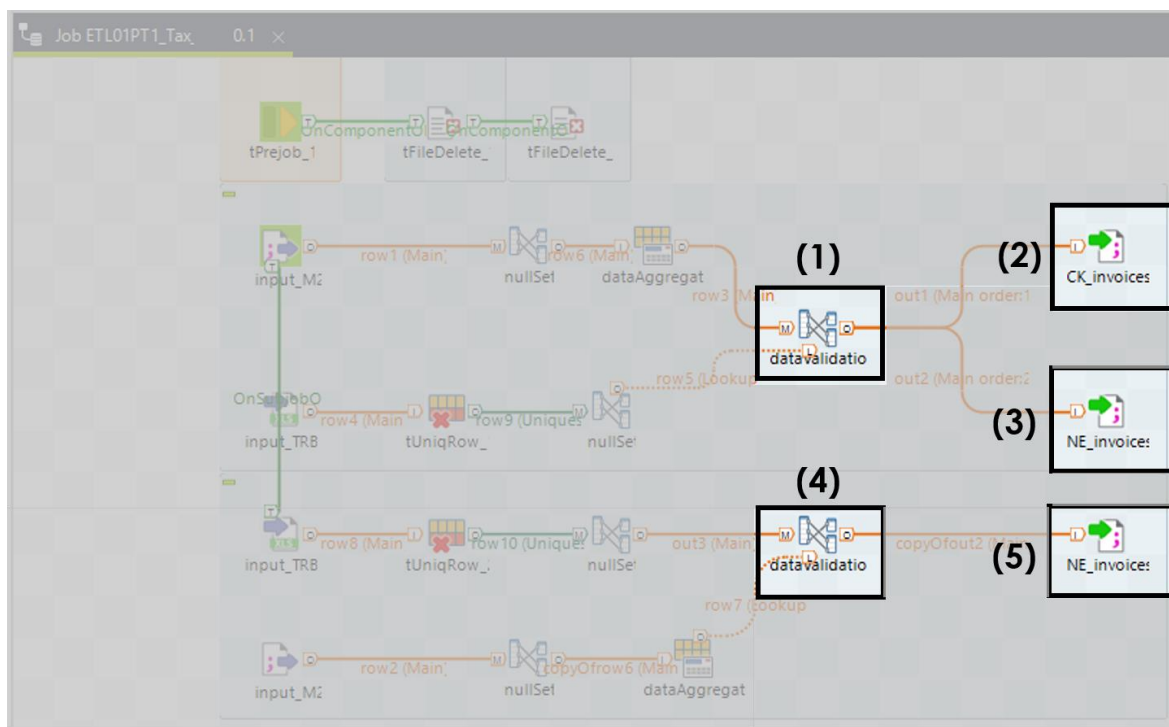
O segundo *tMap* da Figura 3.29 (4) foi utilizado com a seguinte intenção:

- Como as notas fiscais em comum nas duas bases já foram comparadas dentro do primeiro componente, esse segundo serviu para rastreamos quais notas fiscais subiram no sistema externo da SEFAZ, mas não foram encontradas no relatório interno da companhia. Essas notas foram exportadas em um arquivo

3 – Revisão Teórica e Modelagem do Sistema Estudado

chamado “NE_invoices” da Figura 3.29 (5) assim como o primeiro para que pudesse ser incorporado ao primeiro relatório chamado da mesma forma.

Figura 3.29 Validações e *outputs* exportados Fonte: Talend

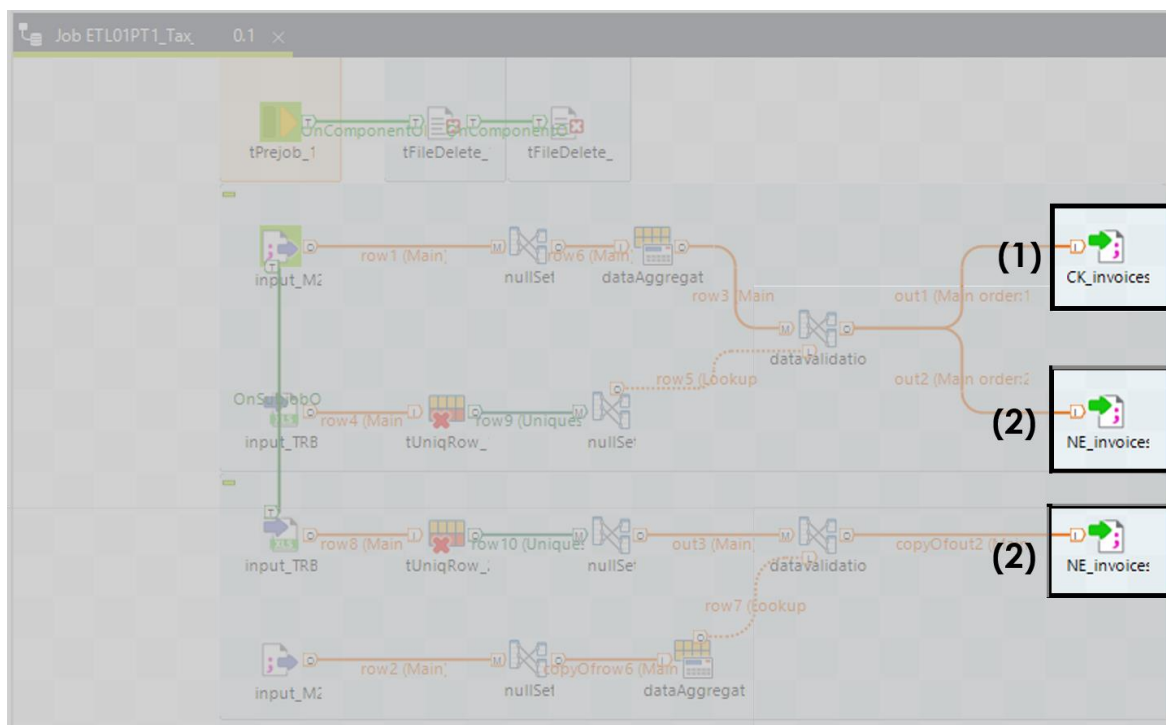


4. RESULTADOS E DISCUSSÕES

Neste capítulo será apresentado o resultado final de todas as tratativas realizadas pela automação, demonstradas no capítulo 3. Esta etapa irá detalhar como devem ser os *outputs* esperados ao final da execução do *job*. Saídas essas que devem ser consumidas pelos analistas da área, substituindo as verificações manuais e insustentáveis que os mesmos realizam antes de obterem o robô atuando junto á operação de rotina da área de negócio.

A princípio, tem-se dois relatórios finais exportados pela automação, o “CK_invoices” – Figura 4.1 (1) - (que contém as notas encontradas em ambas as bases e que passaram pelas verificações de quantidades e valores) e “NE_invoices” – Figura 4.1 (2) - (que é estruturado tanto por notas que foram encontradas na primeira base, mas não foram rastreadas na base dois, quanto por notas fiscais existentes na segunda base, porém não foram encontradas na primeira base de dados).

Figura 4.1 Relatórios finais do *job* destacadas Fonte: Autor



Pode-se observar previamente nas duas bases as notas fiscais rastreadas por chave de acesso, número da nota e série, que são comuns nas duas bases destacadas abaixo:

4 – Resultados e Discussões

Tabela 7 Bases com as notas fiscais em comum destacadas Fonte: Autor

Base 1						Base 2					
invoice_key	num_nf	serie_nf	item_amount	item_price_tax	item_price_wtax	invoice_id	nf_id	serie_id	amount_id	total_value_invoice_tax	total_value_invoice
666796586	23564	3	27	R\$ 1.903.830,60	R\$ 1.813.172,00	666796586	23564	3	27	R\$ 1.903.830,60	R\$ 1.813.172,00
869118612	12587	5	18	R\$ 526.756,65	R\$ 501.673,00	869118612	12587	5	18	R\$ 526.756,65	R\$ 501.673,00
661187273	45879	4	23	R\$ 297.785,25	R\$ 283.605,00	661187273	45879	4	22	R\$ 297.785,25	R\$ 283.605,00
510804431	56987	1	24	R\$ 1.500.703,20	R\$ 1.389.540,00	510804431	56987	1	24	R\$ 1.500.703,20	R\$ 1.389.540,00
352572494	25964	2	7	R\$ 474.723,72	R\$ 439.559,00	352572494	25964	2	7	R\$ 474.723,72	R\$ 439.559,00
517639823	23648	8	8	R\$ 269.457,84	R\$ 249.498,00	517639823	23648	8	8	R\$ 269.457,84	R\$ 249.498,00
873098214	90980	1	65	R\$ 49.282,56	R\$ 45.632,00	588982809	20548	7	32	R\$ 964.199,16	R\$ 892.777,00
432610093	116782	5	32	R\$ 710.521,20	R\$ 657.890,00	171267229	98512	2	27	R\$ 1.073.188,44	R\$ 993.693,00
588982809	20548	7	32	R\$ 964.199,16	R\$ 892.777,00	793449696	12546	6	21	R\$ 179.967,78	R\$ 176.439,00
171267229	98512	2	27	R\$ 1.073.188,44	R\$ 993.693,00	256487961	15628	8	12	R\$ 63.720,00	R\$ 59.000,00
793449696	12546	6	21	R\$ 164.012,94	R\$ 160.797,00	420360517	75013	6	9	R\$ 31.270,32	R\$ 28.954,00
882027439	65824	8	17	R\$ 847.586,34	R\$ 830.967,00	882027439	65824	8	17	R\$ 847.586,34	R\$ 830.967,00
346711096	90854	6	48	R\$ 1.687.465,56	R\$ 1.654.378,00	562245232	20367	9	36	R\$ 626.372,82	R\$ 614.091,00
562245232	20367	9	36	R\$ 626.372,82	R\$ 614.091,00	763564033	56982	4	28	R\$ 929.053,44	R\$ 829.512,00
763564033	56982	4	28	R\$ 929.053,44	R\$ 829.512,00	420954339	89520	2	24	R\$ 1.377.148,64	R\$ 1.229.597,00
420954339	89520	2	24	R\$ 1.377.148,64	R\$ 1.229.597,00	965832130	26996	1	24	R\$ 146.845,44	R\$ 135.968,00
280719603	56321	8	14	R\$ 633.446,24	R\$ 565.577,00	280719603	56321	8	14	R\$ 633.446,24	R\$ 565.577,00
587028655	85224	4	56	R\$ 1.022.925,12	R\$ 913.326,00	587028655	85224	4	74	R\$ 1.347.648,84	R\$ 1.247.823,00

O reporte final esperado das notas encontradas nas duas bases e foram comparadas é o seguinte:

Tabela 8 Reporte final esperado das notas fiscais comuns entre as bases Fonte: Autor

chave_acesso	numero_nf	serie_nf	check_quantidade	check_valor_impostos	check_valor_total
666796586	23564	3	OK	OK	OK
869118612	12587	5	OK	OK	OK
661187273	45879	4	Diferença de quantidade de: 1	OK	OK
510804431	56987	1	OK	OK	OK
352572494	25964	2	OK	OK	OK
517639823	23648	8	OK	OK	OK
588982809	20548	7	OK	OK	OK
171267229	98512	2	OK	OK	OK
793449696	12546	6	OK	Diferença no valor dos impostos em: R\$ 15.954,84	Diferença no valor total da nota de: R\$ 15.642,00
882027439	65824	8	OK	OK	OK
562245232	20367	9	OK	OK	OK
763564033	56982	4	OK	OK	OK
420954339	89520	2	OK	OK	OK
280719603	56321	8	OK	OK	OK
587028655	85224	4	Diferença de quantidade de: 18	Diferença no valor dos impostos em: R\$324.723,72	Diferença no valor total da nota de: R\$ 334.497,00

As três primeiras colunas da Tabela 8, “*chave_acesso*”, “*numero_nf*” e “*serie_nf*”, servem para os analistas que consumirem o relatório, poderem identificar qual a nota fiscal está sendo verificada. Logo em seguida a coluna “*check_quantidade*” faz uma operação de subtração das quantidades fornecidas pelas duas bases em questão e informa se há alguma diferença. Da mesma forma as colunas “*check_valor_impostos*” e “*check_valor_total*” são verificadas e destacadas caso seja encontrado alguma anomalia.

As colunas com “OK” significam que a automação fez as devidas verificações e conseguiu apurar que os valores estão idênticos nas duas bases, assim permitindo o faturamento das notas fiscais no sistema.

Neste sentido, a linha 4 da Tabela 8, contendo a terceira nota fiscal, apresenta uma inconsistência nas quantidades comparadas entre os dois relatórios. Quando isso acontece, é destacado para o analista logo na identificação da nota (três primeiras colunas) que há algum problema com aquela nota fiscal.

Já na linha 10 da Tabela 8, com a nona nota fiscal, as quantidades nos dois relatórios estão corretas, são idênticas. Porém, foi encontrado uma diferença de R\$15.954,84 no valor da nota com os impostos incidentes e de R\$15.642,00 no valor puramente dos materiais da nota fiscal. Dessa forma os analistas podem tomar as devidas tratativas para o fornecedor ou cliente responsável por essa inconsistência.

A linha 16 da Tabela 8, relacionada ao décimo quinto documento fiscal, contém erros em todas as validações, quantidade, valor dos impostos e valor total da nota fiscal. Facilitando assim o entendimento da tratativa manual a ser realizada, fornecendo um direcionamento mais assertivo aos analistas da área.

O segundo reporte final ilustrado na Tabela 9, é composto pelas notas encontradas no relatório interno da companhia, mas que não foram identificadas na base da SEFAZ e vice-versa:

Tabela 9 Reporte final esperado das notas fiscais não encontradas em alguma das bases

Fonte: Autor

chave_acesso_base1	num_nf_base1	serie_nf_base1	qtde_base1	chave_acesso_base2	num_nf_base2	serie_nf_base2	qtde_base2
873098214	90980	1	65	-	-	-	-
432610093	116782	5	32	-	-	-	-
346711096	90854	6	48	-	-	-	-
-	-	-	-	256487961	15628	8	12
-	-	-	-	420360517	75013	6	9
-	-	-	-	965832130	26996	1	24

Dessa forma, os analistas da área de negócio podem consumir esses dois reportes complementares e atuarem de forma direcionada para cada caso.

É importante destacar que esse tipo de automação é altamente escalável, pois, além de obter-se cerca de 98% de eficácia nas suas execuções dentro do projeto realizado na companhia, tinha-se bases com mais de trezentas mil linhas e a performance alcançada chegou a quinhentas linhas por segundo, ou seja, com uma base de trezentas mil linhas, a automação fazia todas as verificações e exportava os relatórios esperados em apenas 10 minutos.

Os ganhos do projeto real, o qual motivou o presente trabalho são bem expressivos e valiosos para a companhia e alguns deles são:

- Ganho de 1,5 FTE's (*Full Time Equivalent*);
- *Saving* de aproximadamente R\$ 105.000,00 por ano;
- *Cost Avoidance* de até R\$ 450.000,00 por ano, provenientes das verificações agora automáticas, que anteriormente não era possível verificar por inteiro devido a volumetria de notas fiscais;
- Aproximadamente 60.000 notas fiscais emitidas contra a companhia por mês, as quais precisavam passar por essas verificações e eram apenas 3 analistas para apurarem os dados.

4.1 CONCLUSÃO DOS RESULTADOS E DISCUSSÕES

Diante dos resultados apresentados, nota-se que automatizar um processo manual e repetitivo é de grande valia para as grandes empresas. Tem-se ainda ganhos não mensurados como, melhora na qualidade de vida dos analistas da área, centralização de informações, mitigação de erros humanos causados pela exaustão e estresse na realização da mesma atividade por muito tempo. Indicadores esses que corroboram o quanto o projeto foi de grande valia.

Portanto, todos os alinhamentos com a área de negócio, entendimentos dos processos rotineiros feitos pela operação da área se mostraram muito importantes para o desenvolvimento da solução, que visto os resultados apresentados acima obteve grande eficácia. Outro objetivo importante era elaborar e desenvolver um projeto com grande valor agregado no âmbito empresarial para que a cultura de automatização em departamentos fiscais seja cada vez mais usual.

5. CONCLUSÕES GERAIS

Grandes empresas necessitam de uma boa saúde financeira para se projetarem como possíveis concorrentes frente aos *players* de referência em seu nicho específico. Indústrias e companhias *tops of mind* de mercado, estão investindo cada vez mais, desde 2014, em projetos de Orçamento Base Zero que visa tornar os gastos cada vez mais eficientes e enxutos dentro de cada empresa.

Um dos aspectos mais importantes para um bom saneamento do negócio é conter uma contabilidade fiscal em dia e nas melhores boas práticas. Muitas companhias negligenciam essa etapa na revisão de seus gastos e custos anuais, porém conforme foi apresentado, em proporções mínimas, pode-se haver grandes perdas relacionadas ao pagamento indevido de impostos e taxaões relacionadas às obrigações fiscais.

Como uma maneira de mitigar esse prejuízo fiscal, foi proposta uma automação baseada na ideia de tratamento e cruzamento de dados através de ferramentas digitalizadas de ETL. Para isso, desenvolveu-se uma lógica para que as bases utilizadas na verificação dos documentos fiscais, fossem importadas e a partir da entrada dos dados, os mesmos fossem tratados, logo depois comparados e confrontados. Podem ser verificados no apêndice em anexo ao presente trabalho.

Após isso, mostrou-se como os analistas da área poderiam atuar de forma integrada à automação. Que se dá em baixar os relatórios e alocá-los nas pastas especificadas no projeto e dessa forma, a automação importa os dados, os cruza e valida, exportando um relatório final indicando quais os valores comparados e divergidos entre uma base de dados e outra, fazendo com que os colaboradores da área possam ter uma atuação mais analítica dentro do processo e não se desgaste com apurações manuais e repetitivas.

Como trabalhos futuros propõe-se explorar soluções ainda mais robustas para diferentes estruturas de contabilidade fiscal. Desenvolver uma automação que consiga trabalhar com o tratamento de informações importadas de banco de dados na nuvem e que

possam ser integradas com API's (*Application Programming Interface*) de sites governamentais para fazer a comparação de mais relatórios de uma vez.

Além disso, propõe-se estudar novas ferramentas de atuação no mesmo sentido do Talend, ao passo que, afim de ampliar a expertise em *softwares* de tratamento e cruzamento dos dados, podendo assim encontrar ferramentas com interfaces diferentes para escopos diversificados no mundo do negócio.

6. REFERÊNCIAS

- [1] B. I. Márcia, "**Manual de orientação e de procedimentos para as organizações contábeis**" em: CRCRS/RS, Conselho Regional de Contabilidade do Rio Grande do Sul, 2010.
- [2] PESSOA, Guilherme. **Entenda a importância da gestão de risco para sua empresa**. 2018. Disponível em: <https://dattos.com.br/entenda-a-importancia-da-gestao-de-risco-para-sua-empresa/>. Acesso em: 12 nov. 2020.
- [3] JIM YONG KIM. World Bank Group. **DOING BUSINESS 2019: training for reform**. Training for Reform. 2019. Disponível em: https://www.doingbusiness.org/content/dam/doingBusiness/media/Annual-Reports/English/DB2019-report_web-version.pdf. Acesso em: 23 nov. 2020.
- [4] KYVYA REVOREDO (Rio de Janeiro). Escritório de Contabilidade no Rio de Janeiro Rj. **Contabilidade Fiscal: Quais Seus Benefícios Para Empresas?** 2017. Disponível em: <https://contabilrio.com.br/escritorio-de-contabilidade-no-rio-de-janeiro-rj-contabil-rio/>. Acesso em: 23 nov. 2020.
- [5] AUTOMATION ANYWHERE (comp.). **Pesquisa global revela as cinco tarefas administrativas “mais odiadas”**. Disponível em: <https://www.automationanywhere.com/br/company/press-room/global-research-reveals-world-s-most-hated-office-tasks>. Acesso em: 24 nov. 2020.
- [6] ATRACTO. **Automação fiscal é eficiente para gestão de documentos?** 2019. Disponível em: <https://auditto.com.br/automacao-fiscal-e-eficiente-para-gestao-de-documentos/>. Acesso em: 28 nov. 2020.
- [7] GRUPO BLB (org.). **Automação fiscal: os robôs na gestão tributária**. 2019. Disponível em: <https://www.blbbrasil.com.br/blog/automacao-fiscal/>. Acesso em: 25 nov. 2020.

-
- [8] GRUPO IMPACTA (org.). **ETL: o que é e qual sua importância na Integração de Dados**. Disponível em: <https://www.impacta.com.br/blog/etl-o-que-e-e-qual-sua-importancia-na-integracao-de-dados-2/>. Acesso em: 22 nov. 2020.
- [9] LOSHIN, David. **ETL O que é e qual sua importância?** Disponível em: https://www.sas.com/pt_br/insights/data-management/o-que-e-etl.html#:~:text=ETL%20%C3%A9%20um%20tipo%20de,para%20construir%20um%20data%20warehouse. Acesso em: 24 nov. 2020.
- [10] IBM CLOUD EDUCATION (org.). **ETL (Extract, Transform, Load)**. 2020. Disponível em: <https://www.ibm.com/cloud/learn/etl>. Acesso em: 26 nov. 2020.
- [11] AHMEDSHUHEL. **Extract, transform, load**. Disponível em: <https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/WIKIPEDI/W121031E.pdf>. Acesso em: 28 nov. 2020.
- [12] PATHAK, Punit. **ETL — Understanding It and Effectively Using It**. 2019. Disponível em: <https://medium.com/hashmapinc/etl-understanding-it-and-effectively-using-it-f827a5b3e54d>. Acesso em: 27 nov. 2020.
- [13] MOTWANI, Rama. **What is ETL Process? What is a data mart?** 2015. Disponível em: <https://ramamotwani.wordpress.com/2015/09/19/what-is-etl-process-what-is-a-data-mart/>. Acesso em: 22 nov. 2020.
- [14] ACHILLEUS. **Joins in Apache Spark — Part 1**. 2019. Disponível em: <https://medium.com/@achilleus/https-medium-com-joins-in-apache-spark-part-1-dabbbf3475690>. Acesso em: 26 nov. 2020.
- [15] PAWEŁ, Ewelina &. **Talend DI Tutorial: Joining data sources with the tMap component**. Disponível em: ACHILLEUS. Joins in Apache Spark — Part 1. 2019. Disponível em: <https://medium.com/@achilleus/https-medium-com-joins-in-apache-spark-part-1-dabbbf3475690>. Acesso em: 26 nov. 2020.. Acesso em: 26 nov. 2020
- [16] HAWKE, Matt Fitzpatrick e Kyle. **A volta do orçamento Base Zero (OBZ)**. 2019. Disponível em: <http://www.bh1.com.br/management/a-volta-do-orcamento-base-zero-obz/>. Acesso em: 21 nov. 2020.
- [17] TALEND INC. (org.). Talend Data Integration products Release Notes. Disponível em: <https://help.talend.com/r/iAD8HwRiMEUI3A09IfKWpg/xBXI9leAYUsWhuwfjQLjCg>. Acesso em: 24 nov. 2020.

7. APÊNDICE A-MODELOS DE SIMULAÇÃO

Esquemas utilizados para as validações efetuadas:

