

UNIVERSIDADE FEDERAL DE SERGIPE
CAMPUS ALBERTO CARVALHO
DEPARTAMENTO DE SISTEMAS DE INFORMAÇÃO

STHEFANIE CARVALHO RODRIGUES OLIVEIRA

PROJETO E IMPLEMENTAÇÃO DE SUBSISTEMA PARA
AGENDAMENTO, EXECUÇÃO E MONITORAMENTO
DE PROCESSOS ETL

ITABAIANA

2015

UNIVERSIDADE FEDERAL DE SERGIPE
CAMPUS ALBERTO CARVALHO
DEPARTAMENTO DE SISTEMAS DE INFORMAÇÃO

STHEFANIE CARVALHO RODRIGUES OLIVEIRA

PROJETO E IMPLEMENTAÇÃO DE SUBSISTEMA PARA
AGENDAMENTO, EXECUÇÃO E MONITORAMENTO
DE PROCESSOS ETL

Trabalho de Conclusão de Curso
submetido ao Departamento de Sistemas
de Informação da Universidade Federal de
Sergipe, como requisito parcial para a
obtenção do título de Bacharel em
Sistemas de Informação.

Orientador: Msc. ANDRÉ VINICIUS RODRIGUES PASSOS NASCIMENTO
Coorientador: Dr. METHANIAS COLAÇO RODRIGUES JÚNIOR

ITABAIANA
2015

Oliveira, Sthefanie Carvalho Rodrigues.

Projeto e Implementação de Subsistema para Agendamento, Execução e Monitoramento de Processos ETL/ Sthefanie Carvalho Rodrigues Oliveira – Itabaiana: UFS, 2014. 80f.

Trabalho de Conclusão de Curso em Bacharel em Sistemas de Informação – Universidade Federal de Sergipe, Curso de Sistemas de Informação, 2014.

1. Data Warehouse.
2. Banco de Dados.
3. Sistemas de Informação.

I. Projeto e Implementação de Subsistema para Agendamento, Execução e Monitoramento de Processos ETL.

STHEFANIE CARVALHO RODRIGUES OLIVEIRA

**PROJETO E IMPLEMENTAÇÃO DE SUBSISTEMA PARA
AGENDAMENTO, EXECUÇÃO E MONITORAMENTO DE
PROCESSOS ETL**

Trabalho de Conclusão de Curso submetido ao corpo docente do Departamento de Sistemas de Informação da Universidade Federal de Sergipe (DSIITA/UFS) como parte dos requisitos para obtenção do grau de Bacharel em Sistemas de Informação.

Itabaiana, 25 de Fevereiro de 2015

BANCA EXAMINADORA:

Prof. André Vinicius Rodrigues Passos Nascimento, Mestre
Orientador
DSIITA/UFS

Prof. Methanias Colaço Rodrigues Júnior, Doutor
Coorientador
DSIITA/UFS

Prof. Eugênio Rubens Cardoso Braz, Doutor
DSIITA/UFS

DEDICATÓRIA

Dedico esse trabalho aos meus maiores educadores, minha mãe Josineide, por seu cuidado e dedicação e a meu pai Ultembergue por sua segurança e perseverança. Ao meu irmão Guto pelo apoio e compreensão. À todos os meus amigos, pelas alegrias, tristezas e dores compartilhas. E a minha prima irmã Camila, por todo incentivo, puxão de orelha e por me trazer até aqui. Obrigada a todos!

AGRADECIMENTOS

Enfim o grande dia chegou, a Formatura. Momento de festejar a vitória alcançada, momento de despedidas e agradecimentos. Quando subirmos ao palco, em trajes de festa e coração batendo mais forte, talvez não sintamos a maravilhosa magia desse grande momento. Mas um dia, quando a saudade bater e as recordações vierem, iremos nos sentir mais completos e mais inteiros, como hoje.

Primeiramente eu agradeço a Deus, sem ele nada na minha vida faria sentido. É graças a Ele que cheguei onde cheguei, é graças a Ele que eu tive forças para lutar, e foi por seu grande amor e misericórdia que eu alcancei mais essa vitória. A minha família eu dedico mais essa glória.

A minha mãe o meu muito obrigado, obrigado por ser meu exemplo, obrigado pelo seu carinho pelo seu amor e paciência durante essa jornada. A meu pai o meu imenso obrigado pela sua dedicação e seu esforço. Que esse seu coração gigante que contagia a todos, seja meu espelho para toda a vida. A meu irmão Guto, agradeço por entender o meu “estresse” e por ser esse exemplo de determinação.

De forma geral meu muito obrigada a toda minha família, tios, primos e avós. Agradeço a todos os meus amigos, em especial a Vanessa. Sabe aquelas amizades que é pro resto da vida?! Ela é uma delas. Rodrigo e Thauane, amigos que fiz durante essa caminhada e que vão morar pra sempre no meu coração. E o que dizer de Leonardo e Charles, com esses eu sei que posso contar pra tudo. Obrigado a todos! Não poderia esquecer de agradecer também a todos os professores que nos acompanharam durante essa caminhada. Em especial ao professor André Vinicius, agradeço por sua orientação e conselhos.

Não poderia esquecer de agradecer as duas pedras preciosas da minha vida. A Lidianny, minha prima do coração, meu obrigado mais que especial por ser essa jóia na minha vida, agradeço por todos os seus conselhos, por ser luz na minha vida e me aproximar mais ainda do amor de Cristo. E Camila, o que dizer de você? Simples! Eu não estaria escrevendo esse texto agora se não fosse por você. Obrigado por ser minha irmã, minha melhor amiga e meu suporte. Te amo muito. Finalizo com o texto de Jeremias 29, versículos 11, 12 e 13: **“Porque sou eu que conheço os planos que tenho para vocês”** diz o Senhor, **“Planos de fazê-los prosperar e não de causar dano, planos de dar a vocês esperança e um futuro.**

Então vocês clamarão a mim, virão orar a mim e eu os ouvirei. Vocês me procurarão e me acharão quando me procurarem de todo o coração.”

OLIVEIRA, Sthefanie Carvalho Rodrigues. **Projeto e implementação de subsistema para agendamento, execução e monitoramento de processos ETL**. 2015. Trabalho de Conclusão de Curso – Curso de Sistemas de Informação, Departamento de Sistemas de Informação, Universidade Federal de Sergipe, Itabaiana, 2015.

RESUMO

Na construção de um ambiente de *Data Warehouse*, uma das etapas mais importantes é o desenvolvimento do sistema ETL (*Extract, Transform, Load*), que é responsável pela extração dos dados nos sistemas operacionais e fontes externas, pelas transformações que esses dados precisam sofrer e pelo povoamento de fatos e dimensões. Embora cruciais, os processos de Extração, Transformação e Carga constituem apenas uma parte do ciclo de vida do desenvolvimento de um sistema ETL. Tão importante quanto os processos é o correto agendamento, a execução precisa e o monitoramento contínuo. Essas atividades são responsabilidade de um subsistema ETL, conhecido como *Scheduler*. Este trabalho apresenta uma proposta de um *Scheduler* para processos ETL. A partir da utilização de metadados, a abordagem utilizada busca simplificar os processos de agendamento, execução e monitoramento dos processos ETL em um ambiente de *Data Warehouse*.

Palavras-chave:

Data Warehouse, Processo ETL, *Scheduler* ETL.

ABSTRACT

During the creation of a data warehouse environment, one of the most important steps is the development of the ETL system (Extract, Transform, Load), which is responsible for the extraction of data in operational systems and external sources, the transformations that these data have to suffer and the load of facts and dimensions. Although crucial, the processes of extraction, transformation and load are only a part of the life cycle of an ETL system development. As important as the process, is the correct scheduling, the precise execution and the continuous monitoring. These activities are responsibility of an ETL subsystem, known as Scheduler. This paper proposes a Scheduler for ETL processes. With the use of metadata, the used approach tries to simplify the processes of scheduling, execution and monitoring of ETL processes in a data warehouse environment.

Keywords:

Data Warehouse, Process ETL, *Scheduler* ETL.

LISTA DE FIGURAS

Figura 1: Arquitetura de um ambiente de Data Warehouse	16
Figura 2: Exemplo de um Esquema Estrela.....	18
Figura 3: Exemplo de Tabela de Fato (Tabela Fato de Venda de Ingresso).	19
Figura 4: Exemplo de Tabela Dimensão (Tabela Dimensão Filme).	20
Figura 5: Exemplo de uma tabela de agregado.....	21
Figura 6: Exemplo do Processo ETL.....	22
Figura 7: Modelo de Casos de Uso da Ferramenta.....	31
Figura 8: Esquema de Dados da Ferramenta - Job	45
Figura 9: Esquema de Dados da Ferramenta – Dependência entre Cargas ETL.....	46
Figura 10: Esquema de Dados da Ferramenta – Agendamento de Jobs.....	47
Figura 11: Esquema de Dados da Ferramenta – Monitoramento de Cargas	48
Figura 12: Esquema de Dados da Ferramenta – Parâmetros	48
Figura 13: Esquema de Dados da Ferramenta – Data de Referência	49
Figura 14: Esquema de Dados da Ferramenta	50
Figura 15: Tela do Menu Principal da Ferramenta.....	51
Figura 16: Tela de Cadastro do Job	52
Figura 17: Tela de Editar Job da Ferramenta	53
Figura 18: Tela de Editar Job da Ferramenta	54
Figura 19: Tela de Cadastro da Carga de Extração da Ferramenta	55
Figura 20: Tela de Edição da Carga de Extração da Ferramenta	56
Figura 21: Tela de Edição da Carga Extração da Ferramenta	57
Figura 22: Tela de Cadastro da carga de Transformação da Ferramenta	58
Figura 23: Tela de Edição da carga de Transformação da Ferramenta	59
Figura 24: Tela de Edição da carga de Transformação da Ferramenta	59
Figura 25: Tela de Cadastro da carga de Povoamento da Ferramenta	60
Figura 26: Tela de Edição da carga de Povoamento da Ferramenta	60
Figura 27: Tela de Edição da carga de Povoamento da Ferramenta	61
Figura 28: Tela de Vinculação entre a Carga de Transformação e Extração da Ferramenta ...	62
Figura 29: Tela de Vinculação entre a Carga de Povoamento e Transformação da Ferramenta	63
Figura 30: Tela de Vinculação entre as Cargas de Povoamento da Ferramenta	63
Figura 31: Tela de Agendamento do Job da Ferramenta.....	64
Figura 32: Tela de Edição do Agendamento do Job da Ferramenta.....	65
Figura 33: Tela de Edição do Agendamento do Job da Ferramenta.....	65
Figura 34: Tela de Cadastro dos Parâmetros da Ferramenta	66
Figura 35: Tela de Edição dos Parâmetros da Ferramenta	67
Figura 36: Tela de Edição dos Parâmetros da Ferramenta	67
Figura 37: Tela de Monitoramento da Carga de Extração da Ferramenta.....	68
Figura 38: Tela de Monitoramento da Carga de Transformação da Ferramenta.....	69
Figura 39: Tela de Monitoramento da Carga de Povoamento da Ferramenta.....	69

LISTA DE QUADROS

Quadro 1: Requisitos Funcionais e Não Funcionais.....	30
Quadro 2: Caso de Uso em Nível de Sistema de Autenticar Usuário	33
Quadro 3: Caso de Uso em Nível de Sistema de Manter Jobs – Cadastrar Job	34
Quadro 4: Caso de Uso em Nível de Sistema de Manter Cargas ETL – Cadastrar carga de Extração.....	34
Quadro 5: Caso de Uso em Nível de Sistema de Manter Cargas ETL – Editar cargas de Extração.....	35
Quadro 6: Caso de Uso a Nível de Sistema de Manter Cargas ETL – Cadastrar Cargas de Transformação.....	36
Quadro 7: Caso de Uso a Nível de Sistema de Manter Cargas ETL – Editar Cargas de Transformação.....	36
Quadro 8: Caso de Uso a Nível de Sistema de Manter Cargas ETL – Cadastrar Cargas de Povoamento	37
Quadro 9: Caso de Uso a Nível de Sistema de Manter Cargas ETL – Editar Carga de Povoamento	37
Quadro 10: Caso de Uso a Nível de Sistema de Relacionamento entre Cargas ETL – Vincular e Desvincular Cargas de Transformação e Extração	38
Quadro 11: Caso de Uso a Nível de Sistema de Relacionamento entre Cargas ETL – Vincular e Desvincular Cargas de Povoamento e Transformação	39
Quadro 12: Caso de Uso a Nível de Sistema de Relacionamento entre Cargas ETL – Vincular e Desvincular Cargas de Povoamento	40
Quadro 13: Caso de Uso a Nível de Sistema de Manter Parâmetros do Sistema –Cadastrar Parâmetro.....	40
Quadro 14: Caso de Uso a Nível de Sistema de Manter Parâmetros do Sistema – Editar Parâmetro.....	41
Quadro 15: Caso de Uso a Nível de Sistema de Agendar Cargas de Extração, Transformação e Povoamento – Cadastrar Agendamento	41
Quadro 16: Caso de Uso a Nível de Sistema de Agendar Cargas de Extração, Transformação e Povoamento – Editar Agendamento	42
Quadro 17: Caso de Uso a Nível de Sistema de Monitorar Jobs e Cargas – Monitorar Cargas de Extração	43
Quadro 18: Quadro 18: Caso de Uso a Nível de Sistema de Monitorar Jobs e Cargas – Monitorar Cargas de Transformação.....	43
Quadro 19: Caso de Uso a Nível de Sistema de Monitorar Jobs e Cargas – Monitorar Cargas de Povoamento	44

SUMÁRIO

1.	INTRODUÇÃO.....	11
1.1	Objetivos.....	12
1.1.1	Geral	12
1.1.2	Objetivos Específicos	13
1.2	Relevância do Projeto	13
1.3	Metodologia	13
1.4	Estrutura do Trabalho	14
2.	REVISÃO BIBLIOGRÁFICA	15
2.1	Ambiente de Data Warehouse	15
2.1.1	Modelagem Dimensional.....	17
2.1.2	Tabela de Fatos.....	18
2.1.3	Dimensões	19
2.1.4	Agregados.....	20
2.2	Sistemas ETL	21
2.2.1	Extração	23
2.2.2	Transformação	24
2.2.3	Carga.....	25
2.3	<i>Scheduler</i> ETL	26
2.3.1	Ferramentas para <i>Scheduling</i>	27
2.3.2	Soluções para <i>Scheduler</i>	27
2.3.3	Dependências de carga	28
2.3.4	Desempenho ETL	28
2.3.5	Vulnerabilidades dos Processos ETL	29
3.	PROJETO DA FERRAMENTA	30
3.1	Requisitos Funcionais e Requisitos não Funcionais	30
3.2	Casos de Uso.....	31
3.2.1	Casos de Uso – Nível de Usuário	31
3.2.2	Caso de Uso – Nível de Sistema.....	32
3.2.2.1	Caso de Uso: Autenticar Usuário.....	33
3.2.2.2	Caso de Uso: Manter Jobs	33
3.2.2.3	Caso de Uso: Manter Cargas ETL	34
3.2.2.4	Caso de Uso: Manter Relacionamento entre Cargas ETL	37
3.2.2.5	Caso de Uso: Manter Parâmetros do Sistema.....	40
3.2.2.6	Caso de Uso: Agendar Cargas de Extração, Transformação e Povoamento ..	41
3.2.2.7	Caso de Uso: Monitorar Jobs e Cargas	42
3.3	Modelagem de Dados	44
3.3.1	Componentes do Esquema de Dados.....	44
4.	FERRAMENTA	51
4.1	Procedimentos.....	70
5.	CONCLUSÃO.....	71
	REFERÊNCIAS	73

1. INTRODUÇÃO

A informação é chave para o sucesso de qualquer negócio. A globalização da economia, a mutação dos mercados e o acirramento da concorrência tornaram a informação o bem mais valioso para as organizações, e estas passaram a tratar seus dados não mais como meros resultados de transações, mas como propulsores para atingir melhores resultados. Com base nisso, a partir dos anos 90 surgiu o termo *Ambiente de Data Warehouse* associado a análises de negócio, crescimento e capacidade de prever novas oportunidades (COLAÇO, 2004).

Data Warehouse (DW) é um banco de dados histórico, separado lógica e fisicamente do ambiente de produção da organização, concebido para armazenar dados extraídos deste ambiente. Os dados antes de serem armazenados são selecionados e integrados tornando dessa forma mais fácil o seu acesso. É uma coleção de dados orientada por assuntos, integrado, variante no tempo, e não volátil, que tem por objetivo dar suporte aos processos de tomada de decisão (COLAÇO, 2004), (INMON, 2005).

O processo de construção de um DW é constituído de várias etapas. Embora existam abordagens diferentes para a construção, como apresentadas em (KIMBALL, 2008) e (INMON, 2005), todo processo de construção de um DW possui uma etapa conhecida como sistema ETL (*Extract, Transform and Load*) ou simplesmente ETL. O sistema ETL é considerado crucial, pois está diretamente relacionado com a qualidade dos dados que são armazenados no DW. É na construção do sistema ETL que encontramos os processos ETL, responsáveis pela extração, transformação e carregamento dos dados em um DW.

Embora cruciais, os processos de Extração, Transformação e Carga constituem apenas uma parte do ciclo de vida do desenvolvimento de um sistema ETL. Tão importante quanto os processos é o correto agendamento, a execução precisa e o monitoramento contínuo. Essas atividades fazem parte do subsistema ETL conhecido como *Scheduler*. O *Scheduler* ETL não pode ser visto como uma mera ferramenta para agendar processos. Sua função é garantir que os processos sejam executados no tempo certo, de acordo as relações e dependências existentes. O *Scheduler* ETL é responsável por garantir a integridade dos dados sendo extraídos, transformados e carregados.

Kimball (KIMBALL, 2004) enumera as características principais de um *Scheduler* ETL e aponta as opções existentes para uma solução de *Scheduler*: (i) Ferramenta ETL Integrada; (ii) *Scheduler* de Terceiros; (iii) Sistema Operacional; (iv) Aplicação Customizada.

A primeira opção só está disponível caso o projeto tenha decidido pela utilização de uma Ferramenta ETL. A utilização de Ferramentas ETL em um projeto de Data Warehouse tem vantagens e desvantagens como apontadas em (KIMBALL, 2004). Apesar das vantagens, muitos projetos não fazem uso de Ferramentas ETL, seja pelo preço de licenças, pela arquitetura complexa de soluções *open source*, ou pela experiência dos programadores na construção de outros sistemas ETL. A segunda opção é desprovida muitas vezes de metadados e estratégias ligadas ao processo ETL. São ferramentas genéricas cuja acomodação no ambiente de *Data Warehouse* pode ser onerosa. Na terceira opção temos a utilização de ferramentas nativas como o *Cron* e o *Windows Task Manager*. São ferramentas confiáveis para o agendamento e execução de tarefas, mas também desprovidas de metadados necessários para o ambiente ETL. A última solução, uma Aplicação Customizada, adotada por muitas empresas que não querem utilizar uma Ferramenta ETL, seria a solução mais indicada. No entanto, a sua construção e manutenção depende de uma equipe experiente. Muitas vezes, essa construção é iniciada por consultores que não permanecem por muito tempo nos projetos, criando, dessa forma, um risco para o ambiente de DW.

O objetivo desse trabalho é propor uma quinta alternativa para a solução de *Scheduler* ETL. Seria criar um *Scheduler* ETL independente, baseado em metadados necessários para a execução dos processos ETL. O *Scheduler* ETL funcionaria como um framework responsável por executar os jobs responsáveis pelos processos de extração, transformação e carga. Um projeto utilizando o *Scheduler* ETL independente, seguiria uma especificação para criar seus jobs, mas ficaria livre da implementação dos processos de agendamento, execução e monitoramento, já embutidos na ferramenta.

1.1 Objetivos

1.1.1 Geral

O objetivo desse trabalho é projetar e implementar um subsistema para agendamento, execução e monitoramento de processos ETL.

1.1.2 Objetivos Específicos

- Realizar revisão bibliográfica;
- Fazer um Levantamento de Requisitos para um *Scheduler* ETL;
- Projetar e construir a ferramenta;
- Utilizar a ferramenta em dois casos práticos;
- Analisar o resultado da utilização da ferramenta.

1.2 Relevância do Projeto

O presente trabalho apresenta uma nova proposta para o tratamento do subsistema *Scheduler* para um sistema ETL. Essa nova abordagem busca simplificar os processos de agendamento, execução e monitoramento dos processos ETL. A abordagem simplificada busca favorecer projetos de Ambientes de Suporte à Decisão que optem por não adquirir ferramentas ETL para seus projetos, seja por limitações de recursos financeiros, humanos, ou tempo necessário para aprendizagem de uma nova ferramenta. A ferramenta desenvolvida também pode contribuir para o ensino em disciplinas que abordem conceitos como Sistemas de Suporte à Decisão, *Data Warehouse* e ETL. O presente trabalho também foi relevante na medida em que abordou diferentes fases de um processo de desenvolvimento de sistemas, exigindo do aluno competências como abstração, identificação de melhores soluções, projeto e implementação.

1.3 Metodologia

Inicialmente, foi realizada a revisão da literatura e a pesquisa bibliográfica sobre o ambiente *Data Warehouse* e Sistema ETL. Após a revisão da literatura foi realizado o levantamento de requisitos para o *Scheduler* independente. Após o projeto e

implementação, a ferramenta foi utilizada com dois casos práticos. Ao final, foram analisados os resultados da utilização da ferramenta.

1.4 Estrutura do Trabalho

O restante do trabalho está estruturado como segue. No capítulo 2 são apresentados os principais artigos com o tema do nosso trabalho. No capítulo 3, são descritas as principais atividades realizadas durante o projeto da Ferramenta, levantamento de requisitos, diagrama de casos de uso e modelo lógico. O capítulo 4 descreve um estudo de caso para apresentação das funcionalidades da ferramenta. Finalmente, no capítulo 5 são apresentadas as conclusões do trabalho.

2. REVISÃO BIBLIOGRÁFICA

O objetivo deste capítulo é dissertar sobre os principais conceitos que estão intimamente relacionados com o tema do trabalho. Nas seções a seguir serão abordados os seguintes conceitos: Ambiente de *Data Warehouse*, *Modelagem Dimensional*, *Sistemas ETL* e *Scheduler ETL*.

2.1 Ambiente de Data Warehouse

O termo *Data Warehouse* foi popularizado por (INMON, 2005) que o considera o coração e o alicerce de todo ambiente de suporte à decisão. Ainda segundo (INMON, 2005), podem-se definir como sendo suas principais características, uma coleção de dados orientada por assunto, integrado, não-volátil, variante no tempo, que dá apoio às decisões da administração. (KIMBALL, 2002) define o *Data Warehouse* como “... Uma cópia dos dados da transação especificamente estruturado para consulta e análise.”. Nesse contexto o DW consiste basicamente numa arquitetura cliente/servidor, que fornece uma visão integrada e abrangente de toda organização, facilitando assim todo o processo de tomada de decisão.

A figura 1 ilustra a arquitetura genérica de um ambiente de *Data Warehouse*.

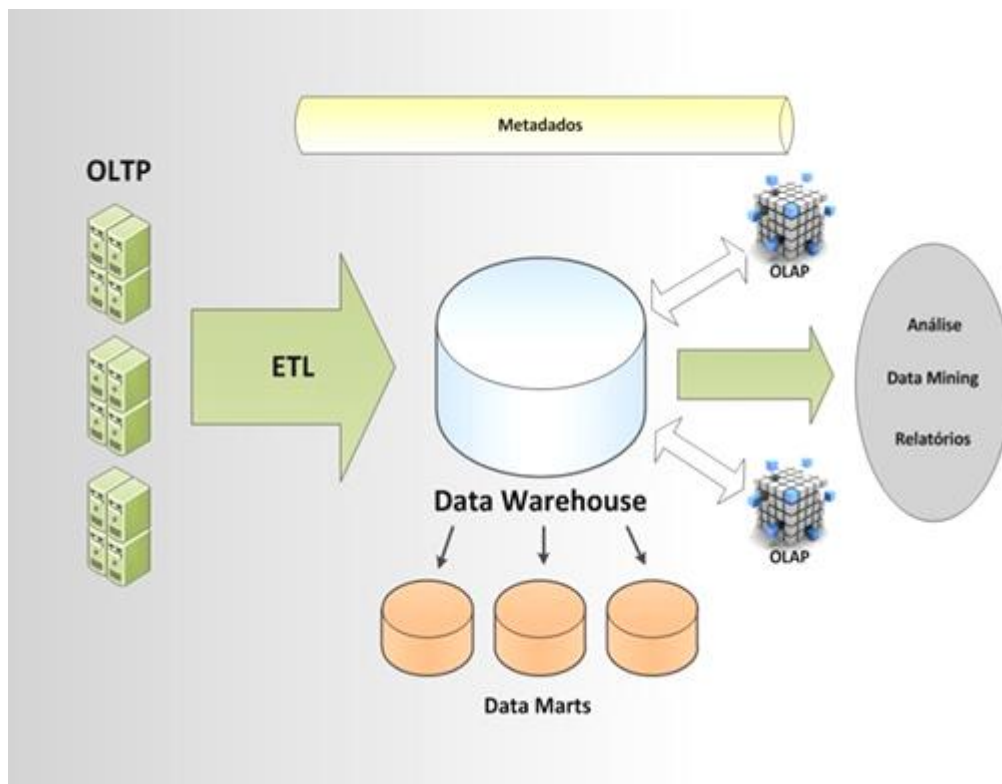


Figura 1: Arquitetura de um ambiente de Data Warehouse

O ambiente **OLTP** (*Online Transaction Processing*, em português, Processamento de Transações em Tempo Real), representa os ambientes transacionais que fornecem os dados fontes que alimentam o DW. Os dados nesse caso são resultado das transações dos sistemas presente no ambiente **OLTP**. Após a extração dos dados, é feita uma filtragem para eliminar os dados que não são necessários. Essa filtragem é realizada na área de **Staging** (Área de Estágio). A área de *Staging* é considerada a área de preparação onde os dados no seu estado bruto sofrem as principais modificações. Integração e limpeza são algumas delas. Logo após, os dados são exportados para o DW. Todo esse refinamento caracteriza os processos de **ETL** (Acrônimo de *Extract-Transform-Load*, que pode ser traduzido como Extração-Transformação-Carga) (KIMBALL, 2002). O processo de ETL é basicamente responsável pela extração, limpeza e carregamento dos dados para o DW, sendo nesse caso as atividades de extração e carga obrigatórias no processo de desenvolvimento de um *Data Warehouse*.

Após todos esses processos, os dados finalmente são carregados para o ambiente de DW. É a partir daí que os dados serão manipulados através de ferramentas que

possam suportar análises de grandes volumes de dados. Essas ferramentas recebem o nome genérico de ferramentas ou tecnologias OLAP. A tecnologia OLAP caracteriza um conjunto de técnicas utilizadas para tratar informações contidas em um *Data Warehouse* (COLAÇO, 2004). Essa classe de ferramenta é utilizada para realizar análises estratégicas, mineração de dados e elaboração de relatórios. As aplicações OLAP são usadas pelos gestores em qualquer nível da organização para lhes permitir análises comparativas que facilitem a tomada de decisões.

Os **Metadados** representam uma espécie de documentação de dados. É a melhor forma de conhecer os dados armazenados em um *Data Warehouse*. O objetivo principal dos Metadados é organizar de forma estruturada todos os dados com o objetivo de diminuir ambiguidades e facilitar o acesso por parte dos usuários. Quando um usuário trabalha com um *Data Warehouse* que não possui Metadados, perde-se um tempo considerável tentando descobrir as informações que existem no *Data Warehouse* assim como as análises que podem ser realizadas (COLAÇO, 2004). Os **Data Marts** representam subconjuntos dos dados contidos em um *Data Warehouse*. Construídos a partir do *Data Warehouse*, apresentam, geralmente, informações voltadas para um Departamento ou Área específica. Além disso, podem apresentar informações resumidas para melhorar o desempenho de análises e consultas.

2.1.1 Modelagem Dimensional

O Diagrama Entidade-Relacionamento (DER) é caracterizado como sendo uma técnica de projeto de banco de dados que visa à eliminação da redundância entre os dados, identificando os dados mais importantes e dados geradores de informação para o processamento de sistemas. Essa técnica é muito eficiente em aplicações transacionais, nas quais os dados são acessados individualmente. Em um ambiente de suporte à decisão, em função das características das consultas e volume de dados, essa técnica tende a criar esquemas difíceis de serem consultados e que apresentam baixo desempenho de consultas.

Dentro desse contexto, (KIMBALL, 2008) define a modelagem dimensional como uma técnica de projeto lógico que busca apresentar os dados em uma estrutura intuitiva e de alto desempenho através da diminuição do grau de normalização. Os esquemas de dados criados com a Modelagem Dimensional recebem o nome de Esquema Estrela ou Esquema em Estrela (KIMBALL, 2008).

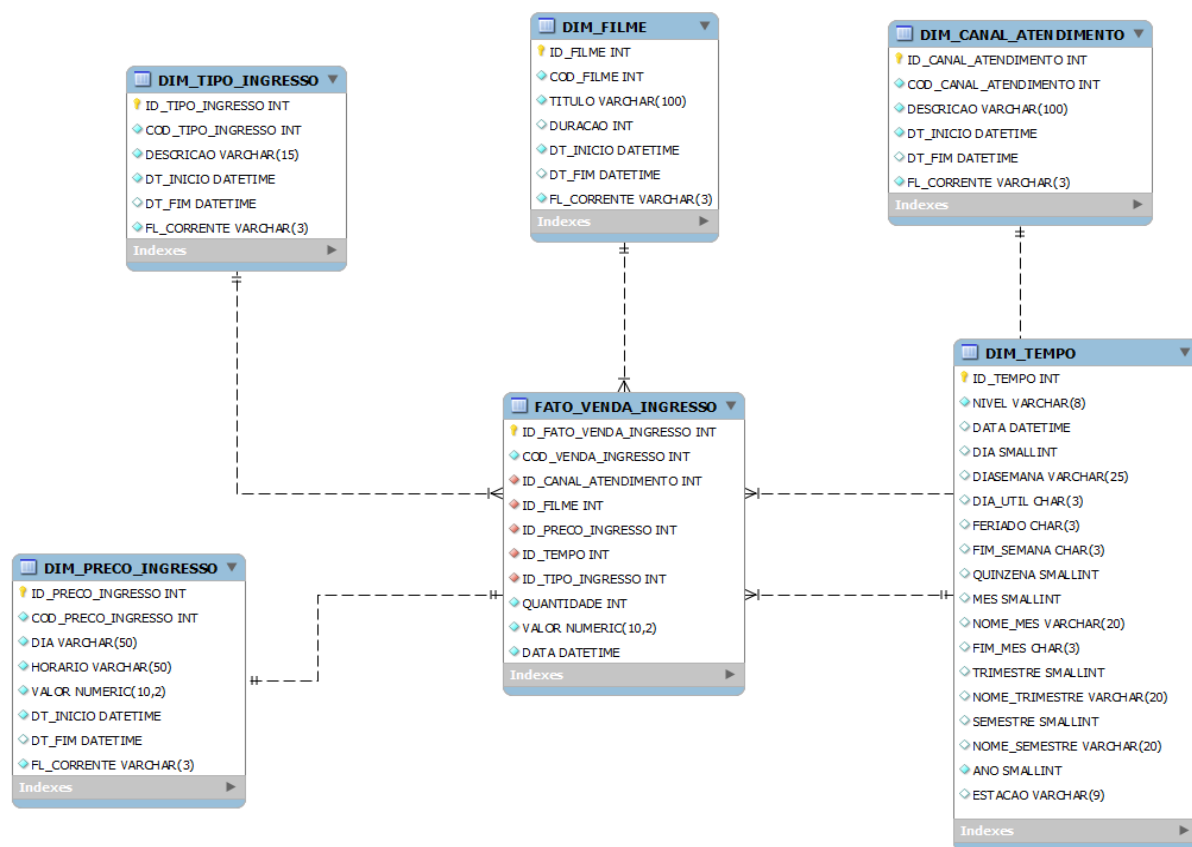


Figura 2: Exemplo de um Esquema Estrela

A figura 2 apresenta um exemplo de um esquema estrela. Este esquema é basicamente formado por uma tabela principal e dominante ao centro, denominada de Tabela de Fatos. Nessa tabela são armazenadas todas as transações e fatos decorrentes dos processos de negócio. A tabela de fatos conecta-se através de junções às demais tabelas auxiliares conhecidas por Dimensões. As dimensões armazenam as características que contextualizam os fatos. Nas próximas seções são descritas as principais características das tabelas de fatos e dimensões.

2.1.2 Tabela de Fatos

De acordo com (KIMBALL, 2008) a tabela de fatos é a principal tabela de um modelo dimensional. Nela são guardadas as medidas numéricas mais importantes do processo de negócio sendo modelado. As tabelas de fato armazenam grande quantidade de dados históricos em função do tempo (COLAÇO, 2004). Todas as tabelas de fatos têm duas ou mais chaves estrangeiras que se conectam às chaves primárias das tabelas de dimensão. Toda tabela de fatos possui uma propriedade conhecida como grão ou granularidade. A granularidade representa o significado de uma linha em uma tabela de fatos. Está associada ao nível de detalhe referente aos registros armazenados. Quanto maior a granularidade, menor o nível de detalhamento.

A figura 3 representa um exemplo de tabela de Fato, Fato Venda de Ingresso. Basicamente, uma tabela de Fato é constituída por dois tipos de atributos: a) Os atributos que representam as ligações com as dimensões; b) Os atributos que representam as medidas do negócio. Na Figura 3 temos os seguintes atributos que representam as ligações com as dimensões: Id_Canal_Atendimento, Id_Filme, Id_Preco_Ingresso, Id_Tempo, Id_Tipo_Ingresso. Os atributos Quantidade, Valor e Data representam as medidas para o fato registrado.

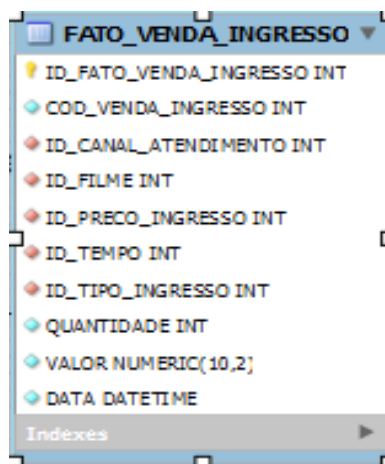


Figura 3: Exemplo de Tabela de Fato (Tabela Fato de Venda de Ingresso).

2.1.3 Dimensões

As dimensões representam tabelas cujos atributos são utilizados como fonte principal de pesquisas, consultas e agrupamentos direcionados à tabela de fatos em um esquema dimensional. Representam a principal fonte de informações e restrições para o processo de análise. Em função disso, os atributos das dimensões devem apresentar valores descritivos e, geralmente, apresentam algum relacionamento hierárquico (IMHOFF; GALEMMO; GEIGER, 2003).

Segundo (COLAÇO, 2004), dimensões representam entidades que evoluem ao longo do tempo. São formadas basicamente por atributos textuais que descrevem como os processos são organizados dentro da dimensão e uma chave primária, na forma de uma *Surrogate Key*.

A figura 4 representa um exemplo de tabela dimensão Filme. Nela, podemos identificar o atributo que representa a chave primária, Id_Filme, e os atributos que descrevem as características do filme sendo modelado, como por exemplo o título do filme e duração.

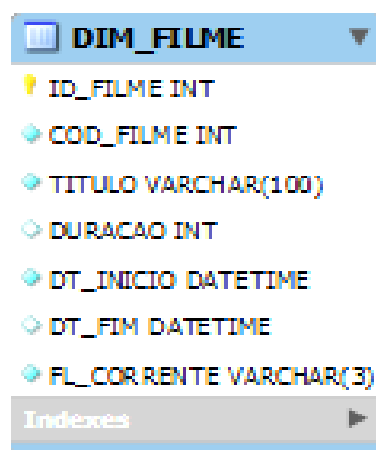


Figura 4: Exemplo de Tabela Dimensão (Tabela Dimensão Filme).

2.1.4 Agregados

A agregação é o processo de resumir dados granulares ou detalhados, e de armazená-los fisicamente em uma tabela de fatos. Como a sumarização dos dados é uma redundância por excelência, os agregados têm um efeito muito significativo sobre o

desempenho das análises e consultas realizadas em um *Data Warehouse*. (KIMBALL, 2008).

O processo de agregação é baseado na criação de novas tabelas com os dados provenientes de tabelas de fatos de menor granularidade. As novas tabelas criadas têm sua granularidade alterada, gerando assim tabelas menores com dados resumidos. Para determinar quais agregados serão criados é preciso analisar quais as principais consultas e resumos que são obtidos da tabela de fatos.

O agregado e as dimensões que o compõe são organizados em um novo esquema estrela no *Data Warehouse*. Um esquema agregado deve sempre fornecer exatamente os mesmos resultados que o esquema base. (ADAMSON, 2006). Na figura 5, temos um exemplo de agregado.

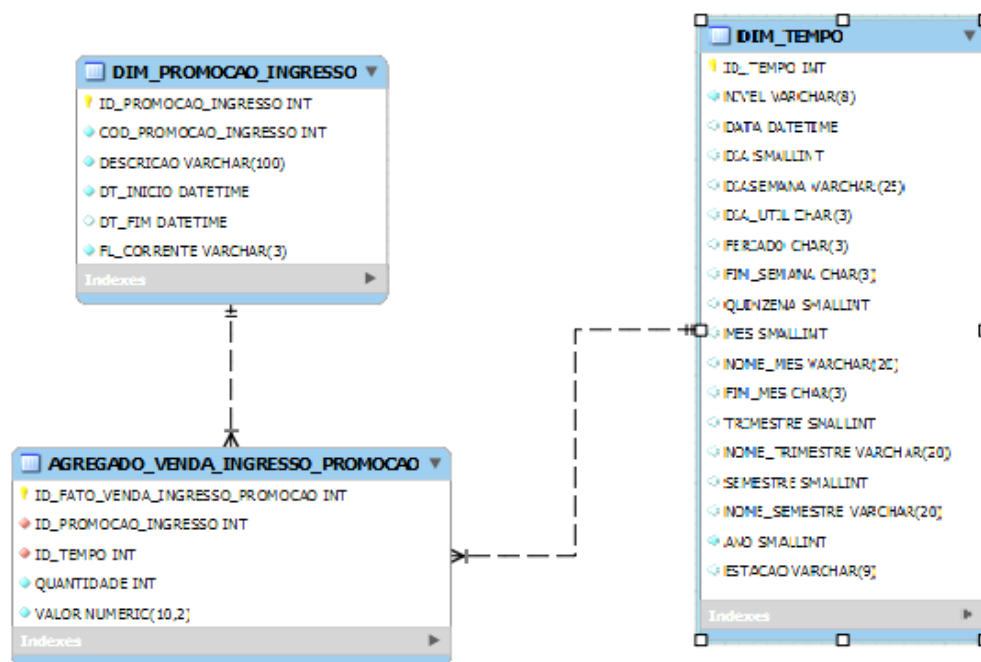


Figura 5: Exemplo de uma tabela de agregado

2.2 Sistemas ETL

Praticamente todos os dados armazenados em um *Data Warehouse* se originam de sistemas externos ao ambiente de Suporte à Decisão. A colocação desses dados e a preparação completa de modo a poder utilizá-los é uma das principais etapas em um *Data Warehouse* (COLAÇO, 2004).

O sistema *Extract-Transform-Load* (ETL) é a camada de integração em um ambiente DW. O Sistema ETL é considerado um dos processos mais críticos na construção de um *Data Warehouse*. Consiste na extração de dados de fontes heterogêneas, na transformação e limpeza necessárias, e na carga dos dados para o DW. Um sistema ETL deve ser responsável, no mínimo, pela execução, controle e oferta dos recursos de monitoramento dos processos de carga no *Data Warehouse* (KIMBALL; CASERTA, 2004), (COLAÇO, 2004). A maior parte do esforço exigido no desenvolvimento de um DW é consumida neste momento. Em geral, cerca de oitenta por cento de todo o esforço de construção é empregado nos processos de ETL (INMON, 1997). A figura 6 representa de forma geral o processo ETL.

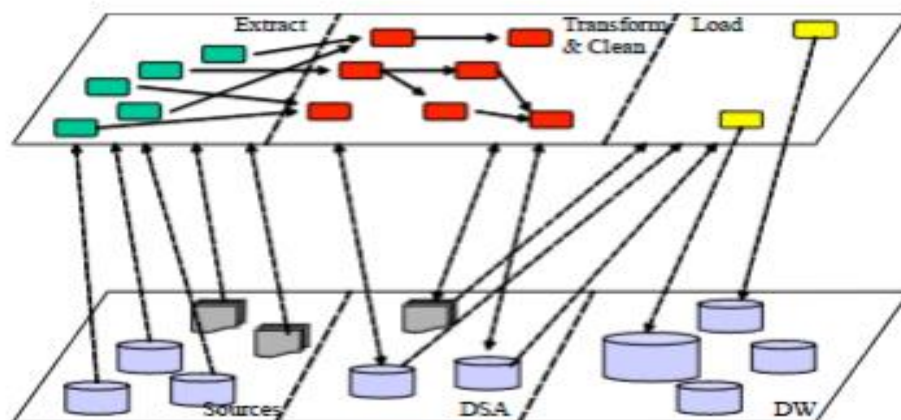


Figura 6: Exemplo do Processo ETL.

Na figura 6, a camada inferior representa o armazenamento de dados que são utilizados em todo o processo. Nessa mesma camada, no lado esquerdo, podemos observar os dados originais provenientes, por exemplo, dos ambientes operacionais das organizações. Como ilustrado na Figura 2, os dados localizados nessas fontes são **extraídos** por meio de rotinas de extração. Posteriormente, esses dados são propagados para a *Data Staging Area* (DSA) onde são transformados e limpos antes de serem carregados para o DW. Na parte direita é representado o *Data Warehouse* (DW), que

tem como objetivo o armazenamento dos dados. O carregamento dos dados no DW é realizado através das atividades de carga representadas na parte superior direita da figura (FERREIRA; MIRANDA; ABELHA; MACHADO).

O ETL é um processo que se divide em três fases: Extração, Transformação e Carga.

2.2.1 Extração

Segundo (KIMBALL, CASERTA, 2004), o verdadeiro desafio na construção de um *Data Warehouse* se dá na integração de todos os dados provenientes de todos os sistemas da organização. E o primeiro passo para o sucesso dessa integração ocorre na extração dos dados. É importante ressaltar que cada organização tem como base dos seus processos diversas fontes distintas. Ao passo que seus negócios evoluem, vão agregando a si, cada vez mais, sistemas que auxiliam na gestão da informação. O principal foco do processo ETL é integrar efetivamente esses diversos sistemas com características variadas, antes de iniciar a construção do processo ETL.

Esta etapa do processo ETL se baseia na busca das informações mais importantes e na reformatação desses dados. Os principais obstáculos que poderiam atrapalhar essa busca seriam a distribuição das origens dos dados, podendo estar em bases com plataformas distintas. Dessa forma, devem ocorrer formas de extração diferentes para cada local. Os mesmos dados que sofreram manipulação por diferentes sistemas podem apresentar características distintas da origem. Nesses casos é preciso extrair esses dados diretamente da fonte que os originou.

Em (KIMBALL, CASERTA, 2004) antes descrever em detalhes sobre as diversas fontes de dados, é necessário explorar o projeto real do documento de mapeamento de dados lógico. Esse documento é necessário para mapear todo o relacionamento entre os dados na sua fonte de origem até o destino. O mapeamento de dados lógico é apresentado na forma de tabela ou planilha. Para que essa análise lógica possa existir é necessário que os sistemas fonte sejam identificados e analisados. A análise do sistema de origem é geralmente dividida em duas grandes fases: A fase de descoberta de dados e a fase de detecção de anomalias.

A fase de descobertas de dados compreende a identificação das principais fontes de origem dos dados. Em organizações onde existem diversas fontes de origem para um

mesmo dado ou informação, é indispensável a definição de um **Sistema de Registro (SOR)** que abrange a fonte originária dos dados. Após essa descoberta, esses sistemas terão seus conteúdos analisados mais profundamente através de um diagrama entidade-relacionamento (COLAÇO, 2004) (KIMBALL, 2008). Após a análise minuciosa dos dados, inicia-se a fase de detecção de anomalias. Nessa fase as anomalias mais encontradas são: valores nulos e atributos de data armazenados em campo texto.

Após todo esse processo de descoberta de dados é realizado a integração dos dados de fontes heterogêneas. Segundo (KIMBALL, CASERTA, 2004) a integração dos dados resulta em uma única fonte de informação, organizada para apoiar o interesse em comum. Em um *Data Warehouse*, a integração ocorre através da implementação de dimensões e fatos.

2.2.2 Transformação

A transformação dos dados é a fase subsequente à sua extração. Esta fase não só transforma os dados, mas também realiza a limpeza dos mesmos. A correção de erros de digitação, a descoberta de violações de integridade, a substituição de caracteres desconhecidos, as padronizações de abreviações podem ser exemplos dessa limpeza (GONÇALVES, 2003). O objetivo principal da limpeza dos dados é a redução dos erros, melhorando suas qualidades e fornecendo dados úteis ao usuário.

No decorrer desse processo muitos problemas podem ser detectados, como por exemplo, a ambiguidade de dados que podem ter a mesma nomenclatura com significados diferentes, ou valores diferentes com o mesmo significado. Existem problemas com a integridade referencial dos dados, onde dados que farão a composição de um fato podem ser negligenciados ou não encontrados na origem no momento de gerar uma dimensão.

Algumas fontes de dados exigirão pouca ou até mesmo nenhuma manipulação nos dados. Outros casos irão requerer alguns determinados tipos de transformação. Segundo KIMBALL (1998), as características mais relevantes para garantir a qualidade dos dados são: a) unicidade, evitando assim duplicações de informação; b) precisão, dados não podem perder suas características originais assim que são carregados para o DW; c) completude, não gerando dados parciais de todo o conjunto relevante às análises e consistência, ou seja, os fatos devem apresentar consistência com as dimensões que o

compõem. É necessário que os dados fiquem em uma forma homogênea para serem carregados no DW.

O usuário ao ter acesso aos dados finais quer a certeza de que aquela é uma fonte confiável de informações. Para isso, o subsistema responsável pela transformação e limpeza deve ser completo na sua detecção, correção e documentação do histórico das transformações realizadas. Deve ser rápido e corretivo no processamento de um grande volume de dados e, principalmente, apresentar práticas transparentes. Essas práticas devem indicar os principais defeitos que prejudicam a qualidade dos dados chamando a atenção para os principais motivos que os tornaram assim. No entanto torna-se inviável que essas práticas sejam realizadas de forma simultânea. Devem ser balanceadas e refletindo assim as principais prioridades de cada uma.

2.2.3 Carga

A última etapa do processo de ETL é a carga. Os dados já extraídos e transformados são carregados para dentro do DW. Basicamente os dados são carregados em dimensões estáticas, dimensões de modificação lenta e fatos integrantes ao modelo do DW. Dimensões estáticas normalmente não oferecem problemas, pois mantém os dados que não sofrem alteração na sua origem e são carregados uma única vez. Já as dimensões de modificação lenta necessitam da verificação em suas fontes e auditorias de cargas (KIMBALL, 1998).

As dimensões estando corretamente carregadas, inicia-se a carga dos fatos. Fatos demandam cuidado na sua carga como o uso das chaves artificiais das dimensões para que se tenha uma integridade referencial e controle de valores nulos. Técnicas para amenizar o processo devido ao grande volume de dados podem ser usadas. Dentre essas técnicas estão: a) A carga incremental dos fatos, que irá carregar apenas dados novos ou alterados; b) A execução do processo em paralelo e em momentos de pouco ou nenhum uso do SGBD; c) A utilização de tabelas auxiliares que serão renomeadas como definitivas ao fim da carga (KIMBALL, 1998).

Os principais desafios de uma etapa de carregamento são para acessar destinos disponíveis e para gravar os dados de resultado (transformados e integrados) nos alvos. (KABIRI, CHIADMI, 2013)

Algumas precauções devem ser tomadas antes de se iniciar a carga dos dados, como por exemplo, o desligamento de índices e referências de integridade, a utilização

de comandos do tipo *truncate* ao invés de *delete*. Ter em mente que no momento da carga alguns dados não serão carregados e deste modo os mecanismos de carga devem dar suporte a auditorias para que possa ser reiniciada no momento em que foi parada. E a possibilidade de manter logs com os dados rejeitados para a avaliação dos motivos pelo qual não foram carregados e assim ajustados para integrarem o conjunto a ser carregado (ALMEIDA, 2006).

2.3 *Scheduler* ETL

O desenvolvimento dos processos de Extração, Transformação e Carga constituem apenas uma parte do ciclo de desenvolvimento de ETL. A parte restante do ciclo de vida é dedicada à execução precisa desses processos. Tão importante quanto os processos ETL é o seu correto agendamento, a execução precisa e o monitoramento contínuo. Essas atividades fazem parte do subsistema ETL conhecido como *Scheduler*. O *Scheduler* ETL é responsável por garantir a integridade dos dados sendo extraídos, transformados e carregados.

Segundo (KIMBALL, CASERTA, 2004), a estratégia de execução ETL cai em duas categorias principais, o agendamento e suporte. O **agendamento** ETL faz muito mais do que organizar as tarefas para serem executadas em um determinado momento. Ele funciona como um mecanismo confiável para gerenciar a implementação física, envolvendo a designação de relacionamentos e dependências entre as tarefas ETL. Os usuários, assim como algumas aplicações, dependem do DW para o fornecimento de informações. Se os dados não forem carregados de uma forma consistente, não terão credibilidade alguma. Para ter a certeza de que os processos ETL são executados e concluídos, o **suporte** funcionará como uma equipe de apoio à produção.

O objetivo de um *Data Warehouse* é, acima de tudo, ser uma fonte confiável de dados, consistente e com o objetivo principal de fornecer dados que auxiliem na tomada de decisão. Para ser um sucesso é necessário que os processos do sistema ETL apresentem três características principais: confiabilidade, disponibilidade e gerenciamento. Os processos ETL precisam ser confiáveis, precisam ser executados de uma forma consistente e sem falhas. Na disponibilidade, o DW, deve ser uma aplicação sempre em execução e disponível para uso. No gerenciamento, os processos ETL devem ter a capacidade de se expandir e evoluir de acordo com a carga e volume de dados.

2.3.1 Ferramentas para *Scheduling*

Esta seção explanará sobre as principais opções disponíveis para automatizar o processo ETL, ou seja, soluções de *Scheduler*. Entretanto é necessário analisar a necessidade de algumas funcionalidades presentes nessas ferramentas antes de escolhê-las.

Em alguns casos, o DW é alimentado por dados provenientes de fontes externas. Esses dados por sua vez, são armazenados em diretórios no sistema. As soluções de *Scheduler* devem ser capazes de reconhecer quando esses tipos de dados chegam ao sistema e iniciar a execução automática. Os Tokens são arquivos criados nos sistemas de arquivos para acionar um evento ETL quando isso acontecer. Além disso, outra característica importante nas soluções de *Scheduler* é a sua capacidade de executar os processos ETL várias vezes ao dia. O sistema de agendamento deve ser capaz de suportar grande número de processos que são executados ao longo do dia em tempo real e a sua execução deve funcionar como um relógio, sem qualquer intervenção humana e sem falhas. Caso ocorra algum problema com os processos, as soluções de ETL devem ter a capacidade de comunicar diferentes grupos ou pessoas, dependendo do trabalho ou do tipo de falha. Essa comunicação pode ocorrer através de mecanismos de notificação ou scripts personalizados.

2.3.2 Soluções para *Scheduler*

Nesta seção, descreveremos quatro opções de solução de *scheduler*. O segredo é escolher uma opção robusta que atenda a todas as necessidades e critérios para o carregamento dos dados no *Data Warehouse*. As quatro opções são: (i) Ferramenta ETL Integrada; (ii) *Scheduler* de Terceiros; (iii) Sistema Operacional; (iv) Aplicação Customizada.

Em relação à ferramenta ETL integrada, de acordo com (KIMBALL, CASERTA, 2004) algumas delas oferecem o mínimo de funcionalidades ou são bastante robustas. Alguns dos principais objetivos em utilizar essas ferramentas incluem o suporte oferecido pelo fornecedor da ferramenta, acelerando dessa forma o processo de resolução de problemas técnicos. Apesar dessa vantagem, muitos projetos não fazem uso de Ferramentas ETL, seja pelo preço de licenças, pela arquitetura complexa de

soluções *open source*, ou pela experiência dos programadores na construção de outros sistemas ETL.

A opção de *Scheduler* de Terceiros possui a desvantagem de ser desprovida muitas vezes de metadados e estratégias ligadas ao processo ETL. São ferramentas genéricas cuja acomodação no ambiente de *Data Warehouse* pode ser onerosa.

Sobre o sistema operacional, não é incomum que os processos ETL sejam executados por uma ferramenta operacional nativa, como *Cron* e o *Windows Task Manager*. Apesar de serem ferramentas confiáveis para o agendamento e execução de tarefas, de acordo com (KIMBALL, CASERTA, 2004), é recomendável uma ferramenta ETL robusta para a construção e manutenção do emprego de ETL. A principal desvantagem nesse caso é a falta de metadados necessários para o ambiente ETL.

A construção de uma opção personalizada seria uma das soluções mais indicadas e adotadas por muitas empresas que não querem utilizar uma ferramenta ETL. No entanto, a sua construção e manutenção dependem de uma equipe experiente. Muitas vezes, essa construção é iniciada por consultores que não permanecem por muito tempo nos projetos, criando, dessa forma, um risco para o ambiente de DW.

2.3.3 Dependências de carga

A principal função do *Scheduler* ETL é garantir que as atividades de agendamento, execução e monitoramento dos processos ETL sejam concretizadas no tempo certo. É importante respeitar o acordo entre as relações e dependências existentes. Essas dependências constituem os metadados. Os metadados são necessários para um correto funcionamento desses processos e a sua definição é um dos aspectos mais importantes para um bom funcionamento ETL. Através deles é possível saber a quantidade de dados carregados, a quantidade de dados rejeitados, o tempo decorrido para a realização desse processo, bem como, os principais ajustes a que devem ser realizados a fim de garantir uma melhor qualidade dos dados.

2.3.4 Desempenho ETL

É fundamental que, tal como ocorre a expansão no *Data Warehouse*, ocorra igualmente a expansão do processo ETL. Isso significa que a solução de *Scheduler* escolhida seja capaz de alcançar o desempenho necessário para processar grandes

volumes de cargas, muito maiores do que o tamanho do seu volume atual. Escalabilidade e desempenho são atributos indispensáveis ao projetar o ETL.

Como foi apresentado anteriormente, o processo ETL é dividido em três etapas: Extração, Transformação e Carga. O processo de Extração de dados nos sistemas de origem pode consumir a maior parte do desempenho do processo ETL. Processos de Extração frequentemente requerem métodos intensivos, tais como, cursores, procedimentos armazenados e subconsultas correlacionadas. Já em relação ao processo de Transformação, o tempo gasto na sua execução é consideravelmente menor do que a Extração e Carga. A maior parte das manipulações reais dos dados é feita na memória, a uma velocidade relativamente alta. Para estimar o tempo exato desse processo, recomenda-se unir o tempo estimado da Extração e da Carga e subtrair do tempo do processo completo. Sobre o processo de Carga, muitos fatores podem afetar a estimativa real do seu desempenho. É importante a certificação de que o ambiente durante o teste corresponda exatamente às condições físicas que existem no ambiente de produção (KIMBALL, CASERTA, 2004).

2.3.5 Vulnerabilidades dos Processos ETL

O principal objetivo de um processo ETL é selecionar dados, transformá-los e carregá-lo no *Data Warehouse*. E a principal finalidade dos desenvolvedores é projetar processos eficientes, resistentes a acidentes e a eventos inesperados durante a realização dessas tarefas. De acordo com (KIMBALL, CASERTA, 2004) o sistema ETL, pode ser organizado tanto de forma horizontal, ou seja, sem dependências com outros fluxos de dados, bem como, de forma vertical onde existe essa dependência.

O sucesso na execução de um processo ETL depende de inúmeros componentes. Como apontado em (KIMBALL, CASERTA, 2004) as principais causas de falhas em um processo são: Falha na rede, falha de banco de dados, falha de disco, falha de memória, falha de qualidade de dados e atualização do sistema sem aviso prévio. A equipe de ETL deve monitorar e avaliar os processos ETL para garantir que eles operem de forma eficiente e que o *Data Warehouse* seja carregado de uma maneira eficaz.

3. PROJETO DA FERRAMENTA

Após a revisão bibliográfica, iniciou-se o projeto da ferramenta, no qual foram analisados e definidos quais seriam seus requisitos funcionais e não funcionais. Além disso foram modelados o digrama de casos de uso e o projeto lógico de dados.

3.1 Requisitos Funcionais e Requisitos não Funcionais

A ferramenta deve conter, como requisitos funcionais, a manutenção de Jobs que consiste basicamente em criar, modificar, remover e consultar conjuntos de cargas. Deve conter também a manutenção de cargas de Extração, Transformação e Carga, onde o usuário poderá cadastrar as mesmas, bem como, alterar, remover ou realizar consultas de cada carga referida. A manutenção da Área de Staging e da Área de Violação que também permite o cadastro de tabelas auxiliares, assim como, a remoção e consultas. A ferramenta também possibilita a definição das dependências entre as cargas, ou seja, uma carga só poderá ser executada após respeitar as devidas permissões à execução da rotina que vier anterior a ela. É possível realizar o agendamento para execução das cargas tanto em conjunto, como também cargas individuais.

Requisitos Funcionais	
R1:	Manter Jobs (Criação, Modificação, Remoção e Consultas)
R2:	Manter Cargas de Extração
R3:	Manter Cargas de Transformação
R4:	Manter Cargas de Povoamento
R5:	Manter relacionamento entre Cargas de Povoamento, Cargas de Extração e Transformação
R6:	Definir Dependência entre Cargas de Povoamento
R7:	Manter Parâmetros do Sistema
R8:	Agendar Cargas de Extração, Transformação e Povoamento
R9:	Monitorar Jobs e Cargas
R10:	Executar Jobs
R11:	Executar Cargas Individuais
Requisitos Não Funcionais	
R1:	O Sistema só deve permitir acesso aos dados por usuários autorizados
R2:	A ferramenta será Desktop
R3:	O Sistema será desenvolvido na linguagem Java.
R4:	O sistema deve poder ser customizado para trabalhar com diferentes SGBDs Relacionais

Quadro 1: Requisitos Funcionais e Não Funcionais

O quadro 1 apresenta uma lista completa dos requisitos funcionais e não funcionais da ferramenta. Além desses requisitos citados anteriormente, a ferramenta foi desenvolvida para ser Desktop, na linguagem Java e utilizar o SQL Server como banco de dados. O sistema deve poder ser customizado para trabalhar com diferentes SGBDs e o seu acesso só deve ser permitido através de usuário e senha.

3.2 Casos de Uso

3.2.1 Casos de Uso – Nível de Usuário

Após a identificação e análise inicial dos requisitos essenciais para a ferramenta, foi modelado o diagrama de casos de uso, que é mostrado na figura 7. Com ele, pode-se verificar quais são as funções do sistema.

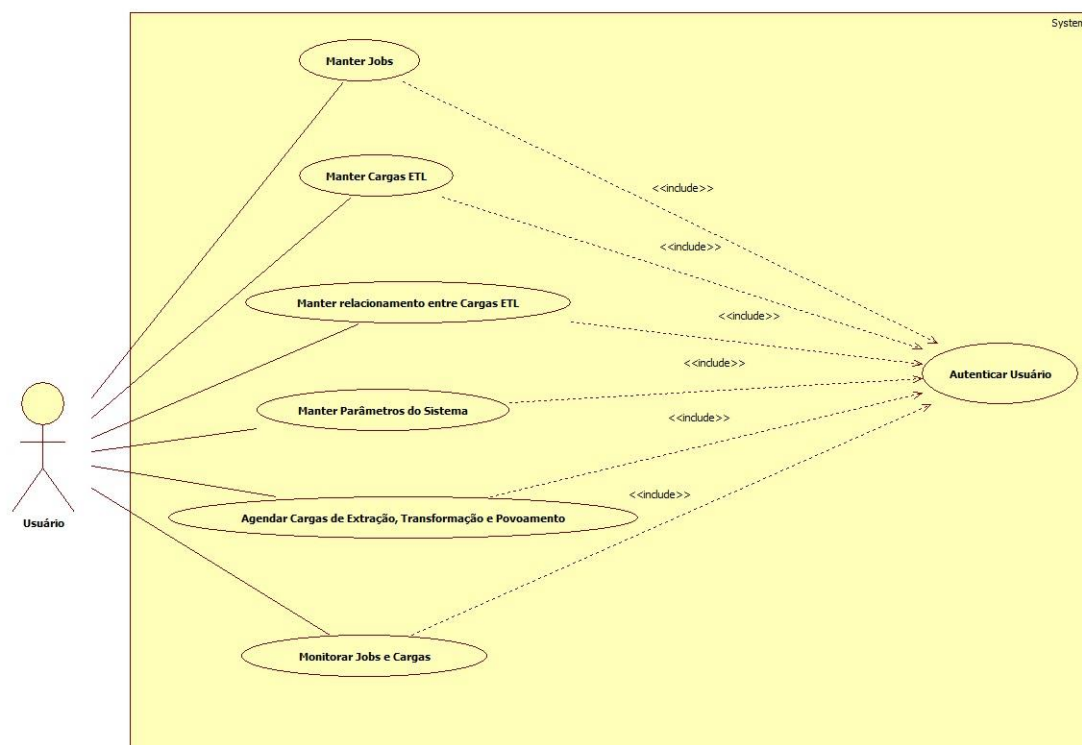


Figura 7: Modelo de Casos de Uso da Ferramenta

O caso de uso Autenticar Usuário valida os dados do usuário, nome de usuário e senha, para permitir seu acesso ao sistema. Somente com a autenticação válida é possível ao usuário acessar as demais funcionalidades da ferramenta. Associa-se ao requisito R1.

O caso de uso Manter Jobs tem como objetivo principal, gerenciar os dados referentes aos Jobs no sistema. Inicialmente o usuário deve acessar o sistema e nele cadastrar o conjunto de cargas. Associado a isso, o usuário também poderá acessar, remover e consultar cada Job cadastrado. Possui o requisito R1 associado.

O caso de uso Manter Cargas ETL, refere-se aos requisitos R2, R3 e R4. Tem o objetivo de permitir o cadastro das cargas de Extração, Transformação e Povoamento. Assim como no caso de uso Manter Jobs, o usuário deve acessar o sistema e cadastrar a referida carga, bem como, remover ou consultar.

O caso de uso Manter relacionamento entre cargas ETL objetiva definir relações de dependências entre as cargas de Extração, Transformação e Povoamento. Todo o processo ETL é dividido em etapas. Para dar seguimento no processo, cada etapa deve ser cumprida por completo, após isso uma nova etapa é iniciada. O sistema permite que essa relação de dependência seja estabelecida. Associa-se aos requisitos R8 e R9.

Agendar cargas de Extração, Transformação e Povoamento é o caso de uso responsável por programar a execução entre as cargas ETL. O sistema permite ao usuário determinar qualquer data específica, dia, mês que deseje que a carga seja executada. Associa-se ao requisito R8.

O caso de uso Monitorar Jobs e Cargas permite ao usuário listar o conteúdo das tabelas que monitoram as cargas de Extração, Transformação e Carga e que trazem o agendamento ou execução de cargas. Associa-se ao requisito R9.

O caso de Executar Jobs é responsável tanto por executar o conjunto de cargas determinados pelo usuário, bem como, cargas individuais caracterizando também o caso de uso Executar Cargas Individuais. Associa-se aos requisitos R10 e R11.

3.2.2 Caso de Uso – Nível de Sistema

Nesta seção apresentaremos os casos de uso em nível de sistema. Nestes são apresentadas o objetivo de cada função, os usuários indicados, pré-condição e pós-

condição para realização do caso de uso. Além disso, os fluxos principais, alternativos e de exceção são detalhados para conhecimento das interações de cada funcionalidade do sistema.

3.2.2.1 Caso de Uso: Autenticar Usuário

Objetivo	Autenticar no sistema para usar funcionalidades disponíveis no sistema.	
Ator	DBA (Administrador de banco de dados) e desenvolvedores.	
Pré-Condição	O Ator precisa estar autenticado no sistema.	
Pós-Condição	Usuário deve estar autorizado para acessar as funcionalidades do sistema.	
Fluxo Principal		
Ação do Ator	Resposta do Sistema	
1. Ator seleciona a opção Conectar do Menu Início.	2. Sistema exibe tela para preenchimento dos dados do usuário.	
3. Ator informa login e senha e clica em Conectar.	4. Sistema verifica preenchimento e validade dos dados.	
	5. Sistema permite o acesso do usuário às funcionalidades.	
Fluxo Alternativo		
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.		
Fluxo Exceção		
Linha 4. Dados não válidos. Sistema exibe mensagem “Usuário/Senha incorretos” e exibe os campos que estão com problemas retornando ao passo 3.		
Linha 4: Campos vazios. Sistema exibe mensagem “Há campos vazios”, retornando ao passo 3.		

Quadro 2: Caso de Uso em Nível de Sistema de Autenticar Usuário

3.2.2.2 Caso de Uso: Manter Jobs

Seção: Cadastrar Job

Objetivo	Cadastrar um novo Job para extração, transformação e povoamento dos dados.		
Ator	DBA (Administrador de banco de dados) e desenvolvedores.		
Pré-Condição	O Ator precisa estar autenticado no sistema.		
Pós-Condição	O Job deve ter sido criado e armazenado numa estrutura de dados específica do sistema.		
Fluxo Principal			
Ação do Ator		Resposta do Sistema	
1. Ator seleciona a opção Cadastrar Job do Menu Jobs.		2. Sistema exibe o formulário para preenchimento dos dados do Job.	

3. Ator informa os dados e clica em Cadastrar.	4. Sistema verifica preenchimento e validade dos dados.
	5. Sistema grava dados informados em meio persistente.
Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	
Linha 4. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.	
Linha 4: Campos vazios. Sistema exibe mensagem “Há campos vazios”, retornando ao passo 3.	

Quadro 3: Caso de Uso em Nível de Sistema de Manter Jobs – Cadastrar Job

3.2.2.3 Caso de Uso: Manter Cargas ETL

Seção: Cadastrar Cargas de Extração

Objetivo	Cadastrar uma nova carga de Extração.
Ator	DBA (Administrador de banco de dados) e desenvolvedores.
Pré-Condição	O Ator precisa estar autenticado no sistema.
Pós-Condição	A carga de Extração deve ter sido cadastrada no sistema.
Fluxo Principal	
Ação do Ator	Resposta do Sistema
1. Ator seleciona a opção Cadastrar no menu Cargas de Extração e sub-menu Cargas.	2. Sistema exibe o formulário para preenchimento dos dados da Carga de Extração.
3. Ator informa os dados e clica em Cadastrar.	4. Sistema verifica preenchimento e validade dos dados.
	5. Sistema grava dados informados em meio persistente.
Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	
Linha 4. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.	
Linha 4: Campos vazios. Sistema exibe mensagem “Há campos vazios”, retornando ao passo 3.	

Quadro 4: Caso de Uso em Nível de Sistema de Manter Cargas ETL – Cadastrar carga de Extração

Seção: Editar Cargas de Extração

Objetivo	Editar os dados da Carga de Extração em meio persistente
-----------------	--

Ator	DBA (Administrador de banco de dados) e desenvolvedores
Pré-Condição	O Ator precisa estar autenticado no sistema
Pós-Condição	A carga de Extração deve ter sido atualizada no Sistema.
Fluxo Principal	
Ação do Ator	Resposta do Sistema
1. Ator seleciona a opção Editar do menu Cargas e Extração e sub-menu Cargas	2. Sistema exibe tela a opção de selecionar a carga que deseja editar.
3. Ator seleciona a carga que deseja editar e clica em Editar.	4. Sistema exibe tela com os dados da carga selecionada.
5. Ator informa os dados que deseja alterar e clica em Editar.	6. Sistema verifica validade dos dados.
	7. Sistema atualiza dados informados em meio persistente.
Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	
Linha 6. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.	

Quadro 5: Caso de Uso em Nível de Sistema de Manter Cargas ETL – Editar cargas de Extração

Seção: Cadastrar Carga de Transformação

Objetivo	Cadastrar uma nova carga de Transformação.
Ator	DBA (Administrador de banco de dados) e desenvolvedores.
Pré-Condição	O Ator precisa estar autenticado no sistema.
Pós-Condição	A carga de Transformação deve ter sido cadastrada no Sistema.
Fluxo Principal	
Ação do Ator	Resposta do Sistema
1. Ator seleciona a opção Cadastrar do menu Cargas de Transformação do sub-menu Cargas.	2. Sistema exibe o formulário para preenchimento dos dados da Carga de Transformação.
3. Ator informa os dados e clica em Cadastrar.	4. Sistema verifica preenchimento e validade dos dados.
	5. Sistema grava dados informados em meio persistente.
Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	
Linha 4. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.	
Linha 4: Campos vazios. Sistema exibe mensagem “Há campos vazios”, retornando ao passo 3.	

Quadro 6: Caso de Uso a Nível de Sistema de Manter Cargas ETL – Cadastrar Cargas de Transformação

Seção: Editar Carga de Transformação

Objetivo	Editar os dados da Carga de Transformação em meio persistente.		
Ator	DBA (Administrador de banco de dados) e desenvolvedores		
Pré-Condição	O Ator precisa estar autenticado no sistema		
Pós-Condição	A carga de Transformação deve ter sido atualizada no Sistema.		
Fluxo Principal			
Ação do Ator		Resposta do Sistema	
1. Ator seleciona a opção Editar do menu Cargas de Transformação do sub-menu Cargas.		2. Sistema exibe tela a opção de selecionar a carga que deseja editar.	
3. Ator seleciona a carga que deseja editar e clica em Editar.		4. Sistema exibe tela com os dados da carga selecionada.	
5. Ator informa os dados que deseja alterar e clica em Editar.		6. Sistema verifica validade dos dados.	
		7. Sistema atualiza dados informados em meio persistente.	
Fluxo Alternativo			
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.			
Fluxo Exceção			
Linha 6. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.			

Quadro 7: Caso de Uso a Nível de Sistema de Manter Cargas ETL – Editar Cargas de Transformação

Seção: Cadastrar Carga de Povoamento

Objetivo	Cadastrar uma nova carga de Povoamento.		
Ator	DBA (Administrador de banco de dados) e desenvolvedores.		
Pré-Condição	O Ator precisa estar autenticado no sistema.		
Pós-Condição	A carga de Povoamento deve ter sido cadastrada no Sistema.		
Fluxo Principal			
Ação do Ator		Resposta do Sistema	
1. Ator seleciona a opção Cadastrar do menu Cargas de Povoamento do sub-menu Cargas.		2. Sistema exibe o formulário para preenchimento dos dados da Carga de Povoamento.	
3. Ator informa os dados e clica em Cadastrar.		4. Sistema verifica preenchimento e validade dos dados.	
		5. Sistema grava dados informados em meio persistente.	
Fluxo Alternativo			
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.			
Fluxo Exceção			

Linha 4. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.
Linha 4: Campos vazios. Sistema exibe mensagem “Há campos vazios”, retornando ao passo 3.

Quadro 8: Caso de Uso a Nível de Sistema de Manter Cargas ETL – Cadastrar Cargas de Povoamento

Seção: Editar Carga de Povoamento

Objetivo	Editar os dados da Carga de Povoamento em meio persistente.
Ator	DBA (Administrador de banco de dados) e desenvolvedores
Pré-Condição	O Ator precisa estar autenticado no sistema
Pós-Condição	A carga de Povoamento deve ter sido atualizada no sistema.
Fluxo Principal	
Ação do Ator	Resposta do Sistema
1. Ator seleciona a opção Editar do menu Cargas de Povoamento do sub-menu Cargas.	2. Sistema exibe tela a opção de selecionar a carga que deseja editar.
3. Ator seleciona a carga que deseja editar e clica em Editar.	4. Sistema exibe tela com os dados da carga selecionada.
5. Ator informa os dados que deseja alterar e clica em Editar.	6. Sistema verifica validade dos dados.
	7. Sistema atualiza dados informados em meio persistente.
Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	
Linha 6. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.	

Quadro 9: Caso de Uso a Nível de Sistema de Manter Cargas ETL – Editar Carga de Povoamento

3.2.2.4 Caso de Uso: Manter Relacionamento entre Cargas ETL

Seção: Vincular e Desvincular Cargas de Transformação e Extração

Objetivo	Definir dependências entre as cargas de Transformação e Extração.
Ator	DBA (Administrador de banco de dados) e desenvolvedores.
Pré-Condição	O Ator precisa estar autenticado no sistema.
Pós-Condição	As cargas de Transformação e Extração devem ter sido cadastradas no Sistema.
Fluxo Principal	

Ação do Ator	Resposta do Sistema
1. Ator seleciona a opção Vincular do menu Cargas de Transformação e Extração do sub-menu Dependências.	2. Sistema exibe na tela a opção para escolher uma carga de Transformação a ser vinculada.
3. Ator seleciona a carga de Transformação que deseja vincular.	4. Sistema exibe na tela uma primeira lista com as cargas de Extração desvinculadas e outra lista com as cargas já vinculadas.
5. Se ator deseja vincular, ele seleciona na primeira lista a carga de Extração que deseja vincular com a carga de Transformação e clica em > ou >>. Se ator deseja desvincular ele seleciona na segunda lista a carga de Extração clica em < ou <<.	6. Sistema verifica validade dos dados.
	7. Sistema salva os dados informados em meio persistente.
Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	
Linha 6. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.	

Quadro 10: Caso de Uso a Nível de Sistema de Relacionamento entre Cargas ETL – Vincular e Desvincular Cargas de Transformação e Extração

Seção: Vincular e Desvincular Cargas de Povoamento e Transformação

Objetivo	Definir dependências entre as cargas de Transformação e Extração.
Ator	DBA (Administrador de banco de dados) e desenvolvedores.
Pré-Condição	O Ator precisa estar autenticado no sistema.
Pós-Condição	As cargas de Povoamento e Transformação devem ter sido cadastradas no Sistema.
Fluxo Principal	
Ação do Ator	Resposta do Sistema
1. Ator seleciona a opção Vincular do menu Cargas de Povoamento e Transformação do sub-menu Dependências.	2. Sistema exibe na tela a opção para escolher uma carga de Povoamento a ser vinculada.
3. Ator seleciona a carga de Povoamento que deseja vincular.	4. Sistema exibe na tela uma primeira lista com as cargas de Transformação desvinculadas e outra lista com as cargas já vinculadas.
5. Se ator deseja vincular, ele seleciona na primeira lista a carga de	6. Sistema verifica validade dos dados.

Transformação que deseja vincular com a carga de Povoamento e clica em > ou >>. Se ator deseja desvincular ele seleciona na segunda lista a carga de Transformação clica em < ou <<.	
	7. Sistema salva os dados informados em meio persistente.
Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	
Linha 6. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.	

Quadro 11: Caso de Uso a Nível de Sistema de Relacionamento entre Cargas ETL – Vincular e Desvincular Cargas de Povoamento e Transformação

Seção: Vincular e Desvincular Cargas de Povoamento

Objetivo	Definir dependências entre as cargas de Transformação e Extração.
Ator	DBA (Administrador de banco de dados) e desenvolvedores.
Pré-Condição	O Ator precisa estar autenticado no sistema.
Pós-Condição	As cargas de Povoamento devem ter sido cadastradas no Sistema.
Fluxo Principal	
Ação do Ator	Resposta do Sistema
1. Ator seleciona a opção Vincular do menu Cargas de Povoamento do sub-menu Dependências.	2. Sistema exibe na tela a opção para escolher uma carga de Povoamento a ser vinculada.
3. Ator seleciona a carga de Povoamento que deseja vincular.	4. Sistema exibe na tela uma primeira lista com as cargas de Povoamento desvinculadas e outra lista com as cargas já vinculadas.
5. Se ator deseja vincular, ele seleciona na primeira lista a carga de Povoamento que deseja vincular com a carga de Povoamento e clica em > ou >>.	6. Sistema verifica validade dos dados.
Se ator deseja desvincular ele seleciona na segunda lista a carga de Povoamento clica em < ou <<.	
	7. Sistema salva os dados informados em meio persistente.
Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	

Linha 6. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.

Quadro 12: Caso de Uso a Nível de Sistema de Relacionamento entre Cargas ETL – Vincular e Desvincular Cargas de Povoamento

3.2.2.5 Caso de Uso: Manter Parâmetros do Sistema

Seção: Cadastrar Parâmetro

Objetivo	Cadastrar uma nova Parâmetro.
Ator	DBA (Administrador de banco de dados) e desenvolvedores.
Pré-Condição	O Ator precisa estar autenticado no sistema.
Pós-Condição	O parâmetro deve ter sido cadastrado no Sistema.
Fluxo Principal	
Ação do Ator	Resposta do Sistema
1. Ator seleciona a opção Cadastrar Parâmetro do menu Parâmetros do sub-menu Início.	2. Sistema exibe o formulário para preenchimento do Parâmetro e do Valor.
3. Ator informa os dados e clica em Cadastrar.	4. Sistema verifica preenchimento e validade dos dados.
	5. Sistema grava dados informados em meio persistente.
Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	
Linha 4. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.	
Linha 4: Campos vazios. Sistema exibe mensagem “Há campos vazios”, retornando ao passo 3.	

Quadro 13: Caso de Uso a Nível de Sistema de Manter Parâmetros do Sistema – Cadastrar Parâmetro

Seção: Editar Parâmetros do Sistema

Objetivo	Editar os dados dos Parâmetros no Sistema.
Ator	DBA (Administrador de banco de dados) e desenvolvedores
Pré-Condição	O Ator precisa estar autenticado no sistema
Pós-Condição	Os Parâmetros devem ter sido atualizados no sistema.
Fluxo Principal	
Ação do Ator	Resposta do Sistema
1. Ator seleciona a opção Editar do menu Parâmetros do sub-menu	2. Sistema exibe tela a opção editar os campos referente ao Parâmetro e Valor.

Início.	
3. Ator altera os valores que deseja editar e clica em Editar.	4. Sistema exibe tela com os novos dados editados.
	5. Sistema verifica validade dos dados.
	6. Sistema atualiza dados informados em meio persistente.
Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	
Linha 6. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.	

Quadro 14: Caso de Uso a Nível de Sistema de Manter Parâmetros do Sistema – Editar Parâmetro

3.2.2.6 Caso de Uso: Agendar Cargas de Extração, Transformação e Povoamento

Seção: Cadastrar Agendamento

Objetivo	Realizar o cadastro do agendamento das Cargas ETL.
Ator	DBA (Administrador de banco de dados) e desenvolvedores.
Pré-Condição	O Ator precisa estar autenticado no sistema.
Pós-Condição	As cargas de Extração, Transformação e Povoamento devem ter sido cadastradas no Sistema.
Fluxo Principal	
Ação do Ator	Resposta do Sistema
1. Ator seleciona a opção Cadastrar Agendamento do menu Agendamento do sub-menu Início.	2. Sistema exibe na tela as opções de selecionar o Job e as demais opções referentes ao agendamento.
3. Ator seleciona o Job e preenche as demais informações e clica em Cadastrar.	4. Sistema exibe tela com os dados do Agendamento.
	5. Sistema verifica validade dos dados.
	6. Sistema atualiza dados informados em meio persistente.
Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	
Linha 6. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.	

Quadro 15: Caso de Uso a Nível de Sistema de Agendar Cargas de Extração, Transformação e Povoamento – Cadastrar Agendamento

Seção: Editar Agendamento

Objetivo	Editar os dados do Agendamento dos Jobs em meio persistente	
Ator	DBA (Administrador de banco de dados) e desenvolvedores	
Pré-Condição	O Ator precisa estar autenticado no sistema	
Pós-Condição	O Job deve ter sido agendado e atualizado no Sistema.	
Fluxo Principal		
Ação do Ator		Resposta do Sistema
1. Ator seleciona a opção Editar Agendamento do menu Agendamento e sub-menu Início.		2. Sistema exibe tela a opção de selecionar a o Job que deseja editar.
3. Ator seleciona o Job que deseja editar e clica em Editar.		4. Sistema exibe tela com os dados do agendamento do Job.
5. Ator informa os dados que deseja alterar e clica em Editar.		6. Sistema verifica validade dos dados.
		7. Sistema atualiza dados informados em meio persistente.
Fluxo Alternativo		
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.		
Fluxo Exceção		
Linha 6. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.		

Quadro 16: Caso de Uso a Nível de Sistema de Agendar Cargas de Extração, Transformação e Povoamento – Editar Agendamento

3.2.2.7 Caso de Uso: Monitorar Jobs e Cargas

Seção: Monitorar Cargas de Extração

Objetivo	Acompanhar o conteúdo do agendamento e execução da Carga de Extração.	
Ator	DBA (Administrador de banco de dados) e desenvolvedores	
Pré-Condição	O Ator precisa estar autenticado no sistema	
Pós-Condição	A carga de Extração deve ter sido agendada e atualizada no Sistema.	
Fluxo Principal		
Ação do Ator		Resposta do Sistema
1. Ator seleciona a opção Monitorar Extração do menu Monitoramento e sub-menu Início.		2. Sistema exibe tela a opção de selecionar a data de referência que deseja verificar.
3. Ator seleciona a data de referência.		4. Sistema verifica validade dos dados.
		5. Sistema exibe a tela todos os dados referente ao agendamento e execução da carga.

Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	
Linha 6. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.	

Quadro 17: Caso de Uso a Nível de Sistema de Monitorar Jobs e Cargas – Monitorar Cargas de Extração

Seção: Monitorar Cargas de Transformação

Objetivo	Acompanhar o conteúdo do agendamento e execução da Carga de Transformação.	
Ator	DBA (Administrador de banco de dados) e desenvolvedores	
Pré-Condição	O Ator precisa estar autenticado no sistema	
Pós-Condição	A carga de Transformação deve ter sido agendada e atualizada no Sistema.	
Fluxo Principal		
Ação do Ator		Resposta do Sistema
1. Ator seleciona a opção Monitorar Transformação do menu Monitoramento e sub-menu Início.		2. Sistema exibe tela a opção de selecionar a data de referência que deseja verificar.
3. Ator seleciona a data de referência.		4. Sistema verifica validade dos dados.
		5. Sistema exibe a tela todos os dados referente ao agendamento e execução da carga.
Fluxo Alternativo		
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.		
Fluxo Exceção		
Linha 6. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.		

Quadro 18: Quadro 18: Caso de Uso a Nível de Sistema de Monitorar Jobs e Cargas – Monitorar Cargas de Transformação

Seção: Monitorar Cargas de Povoamento

Objetivo	Acompanhar o conteúdo do agendamento e execução da Carga de Povoamento.
Ator	DBA (Administrador de banco de dados) e desenvolvedores
Pré-Condição	O Ator precisa estar autenticado no sistema

Pós-Condição	A carga de Povoamento deve ter sido agendada e atualizada no Sistema.
Fluxo Principal	
Ação do Ator	Resposta do Sistema
1. Ator seleciona a opção Monitorar Povoamento do menu Monitoramento e sub-menu Início.	2. Sistema exibe tela a opção de selecionar a data de referência que deseja verificar.
3. Ator seleciona a data de referência.	4. Sistema verifica validade dos dados.
	5. Sistema exibe a tela todos os dados referente ao agendamento e execução da carga.
Fluxo Alternativo	
Linha 3: Ator clica em Fechar. Retorna ao Menu Principal.	
Fluxo Exceção	
Linha 6. Dados não válidos. Sistema exibe mensagem “Dados inválidos” e exibe os campos que estão com problemas retornando ao passo 3.	

Quadro 19: Caso de Uso a Nível de Sistema de Monitorar Jobs e Cargas – Monitorar Cargas de Povoamento

3.3 Modelagem de Dados

Após a criação dos casos de uso da ferramenta, foi modelado o projeto lógico do Sistema. A figura 14, no final da seção, mostra a modelagem de dados completa da ferramenta. Nesta seção detalharemos as tabelas do esquema.

3.3.1 Componentes do Esquema de Dados

Caracterização de Jobs: Todas as informações sobre os Jobs cadastrados são persistidas na tabela TB_JOB. Cada Job com seu identificador, descrição e data de criação. Cada Job pode possuir um conjunto de cargas relacionadas, responsáveis pelos processos de Extração, Transformação e Povoamento dos dados. Esse processo caracteriza o processo ETL. Cada carga é armazenada nas respectivas tabelas, TB_CARGA_EXTRACAO, TB_CARGA_TRANSFORMACAO e TB_CARGA_POVOAMENTO.

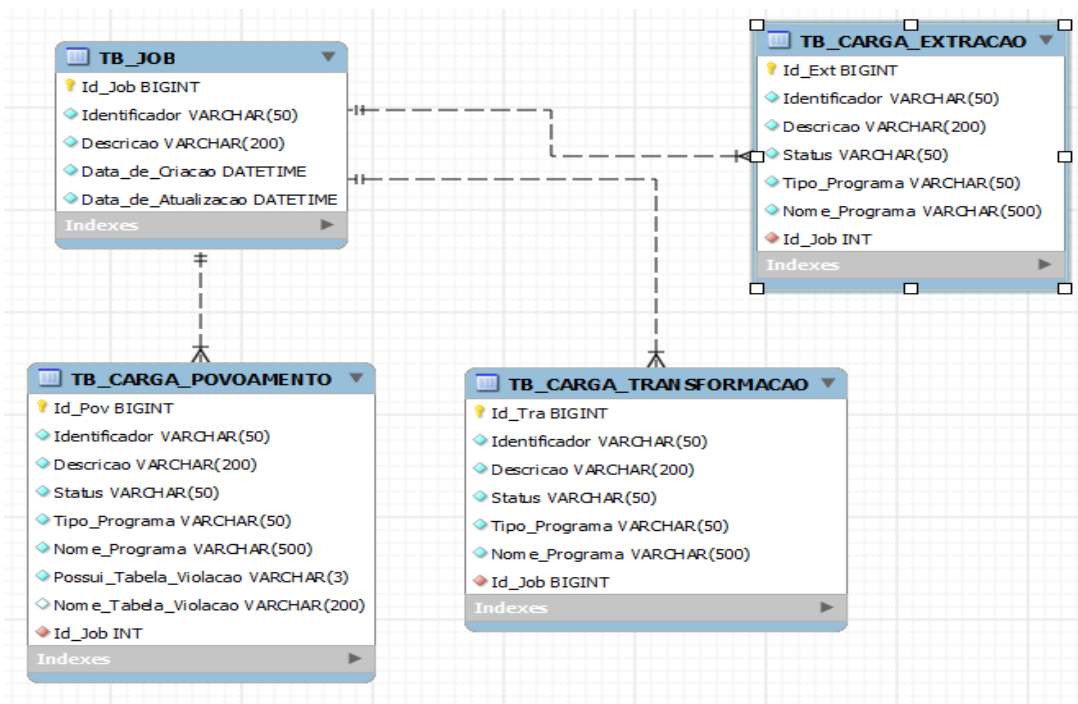


Figura 8: Esquema de Dados da Ferramenta - Job

Caracterização da Carga de Extração: A carga de Extração é responsável basicamente por extrair os dados dos sistemas de origem para a Área de Staging. A TB_CARGA_EXTRACAO tem a finalidade de armazenar essa carga que possui identificador, descrição, status, o tipo do programa e nome do programa. No processo da extração de dados existe uma relação de dependência entre o processo de transformação. A transformação é responsável pela limpeza e integração dos dados em um só padrão. Antes dos dados sofrerem esse processo, eles precisam ser totalmente extraídos das suas fontes. A tabela TB_DEPENDENCIA_TRANSFORMACAO_EXTRACAO é responsável por estabelecer essa relação no sistema.

Caracterização da Carga de Transformação: A carga de Transformação é responsável pela limpeza e tratamento dos dados. Consistem em padroniza-los em relação ao tamanho, tipo, substituição de caracteres estranhos, erros de digitação, duplicidade entre outros. A tabela TB_CARGA_TRANSFORMACAO tem a finalidade de armazenar essa carga que possui identificado, descrição, status, o tipo do programa e nome do programa. A relação de dependência nesse caso, existe entre a carga de povoamento. Antes dos dados migrarem para o DW eles precisam passar por todo esse processo de limpeza e integração, facilitando assim sua atualização no sistema final. A

TB_DEPENDENCIA_POVOAMENTO_TRANSFORMACAO é responsável por estabelecer essa relação no sistema.

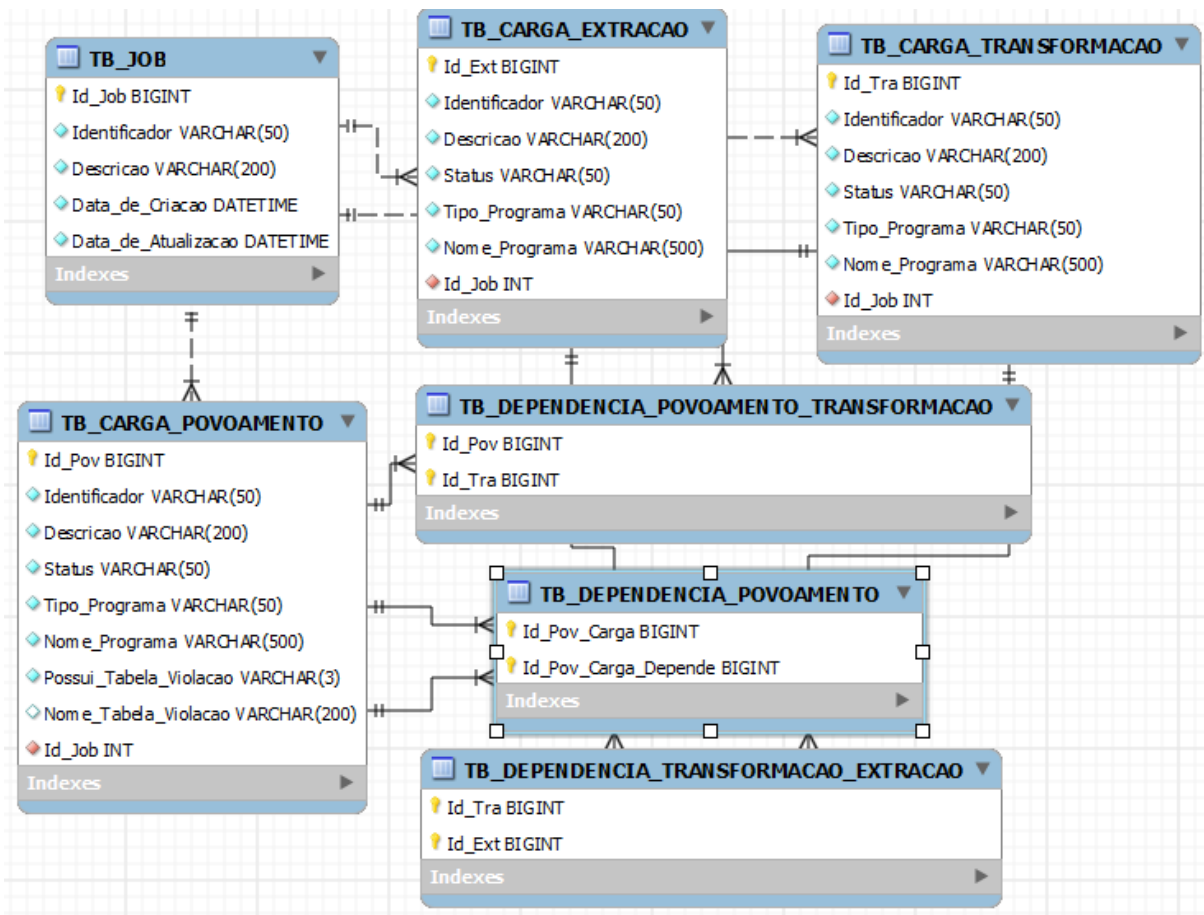


Figura 9: Esquema de Dados da Ferramenta – Dependência entre Cargas ETL

Caracterização do Agendamento: A tabela TB_AGENDAMENTO_JOB permite ao sistema agendar o Job cadastrado conforme as necessidades do usuário. O agendamento pode ser feito através dos campos, dia da semana, dia do mês, dia específico, hora inicial e hora final. Os dias da semana são armazenados na tabela TB_DIAS_DA_SEMANA. Caso a execução do Job seja semanalmente, ou mais de um dia na semana, os dados serão inseridos na tabela TB_AGENDAMENTO_DIAS_DA_SEMANA. A tabela TB_TIPO_FREQUENCIA é usada para definir qual o tipo de agendamento o usuário escolherá.

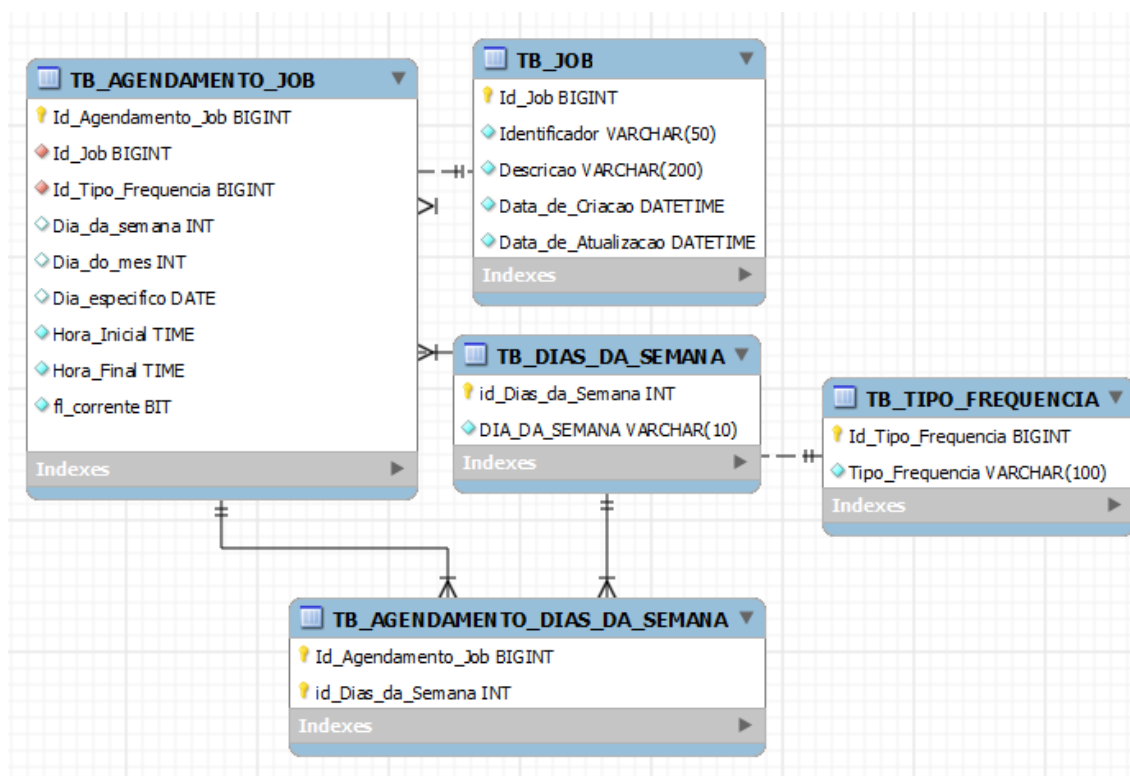


Figura 10: Esquema de Dados da Ferramenta – Agendamento de Jobs

Caracterização de Log: Todas as informações sobre o conteúdo das tabelas de carga em relação a agendamento e execuções, podem ser listadas através do módulo de monitoramento dessas tabelas. Esse monitoramento ocorre através das tabelas de LOG (Figura 11). A TB_LOG_CARGAS_DE_EXTRACAO está relacionada com a tabela TB_CARGA_EXTRACAO, a tabela TB_LOG_CARGAS_TRANSFORMACAO com a tabela TB_CARGA_TRANSFORMACAO, bem como, a tabela TB_LOG_CARGAS_POVOAMENTO com a tabela TB_CARGAS_POVOAMENTO. Os dados a serem listados são descrição, data de referência, data do início do processamento, data do fim do processamento e situação da carga.

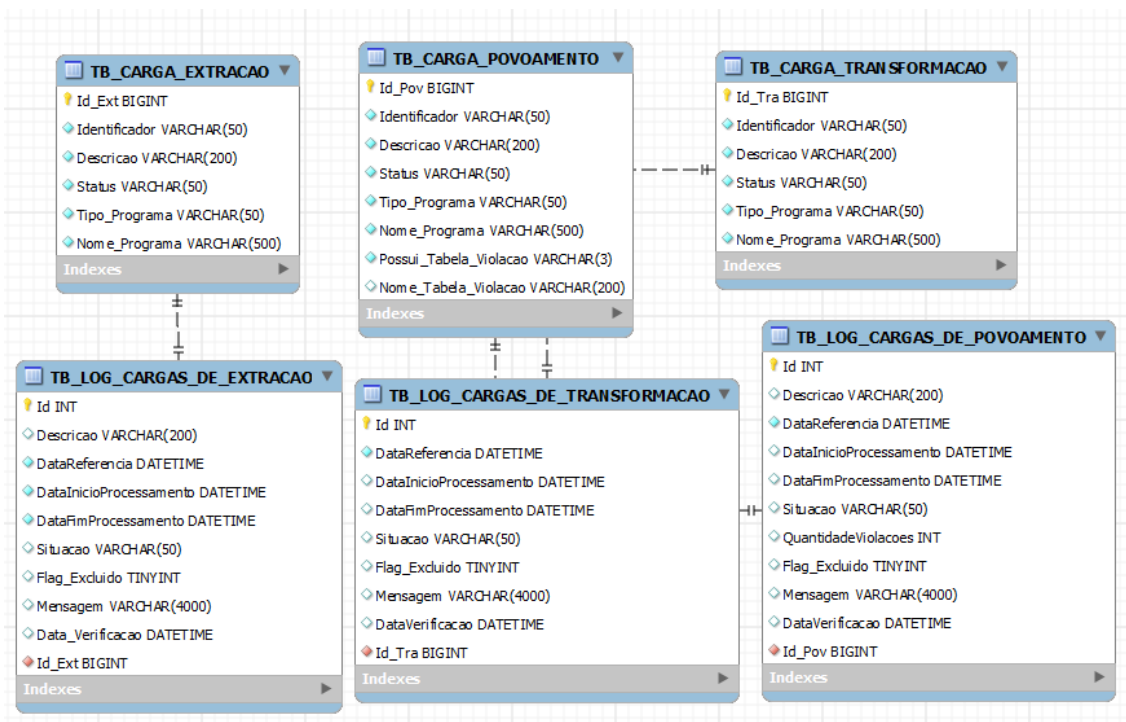


Figura 11: Esquema de Dados da Ferramenta – Monitoramento de Cargas

Caracterização de Parâmetro: A tabela TB_PARAMETRO (Figura 12) armazena quaisquer valores que possam ter alguma utilidade pelo sistema nos campos Parâmetro e Valor. Sejam informações de endereço de email, servidor de email, entre outros.

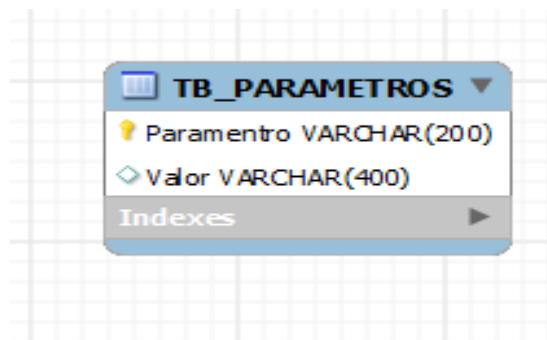


Figura 12: Esquema de Dados da Ferramenta – Parâmetros

Caracterização de Data Referência: O objetivo da tabela TB_DATA_REFERENCIA (Figura 13) é identificar a Data de Referência para a qual as cargas serão executadas. Ela será usada nos procedimentos criados para agendar cargas. Vai existir apenas uma linha nessa tabela.

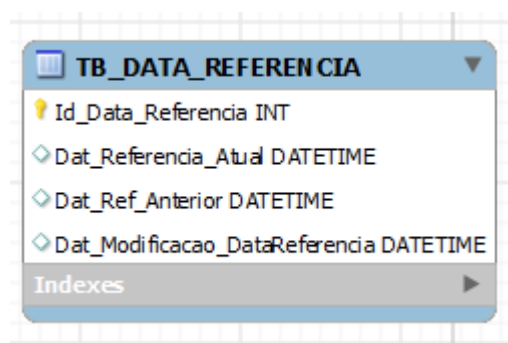


Figura 13: Esquema de Dados da Ferramenta – Data de Referência

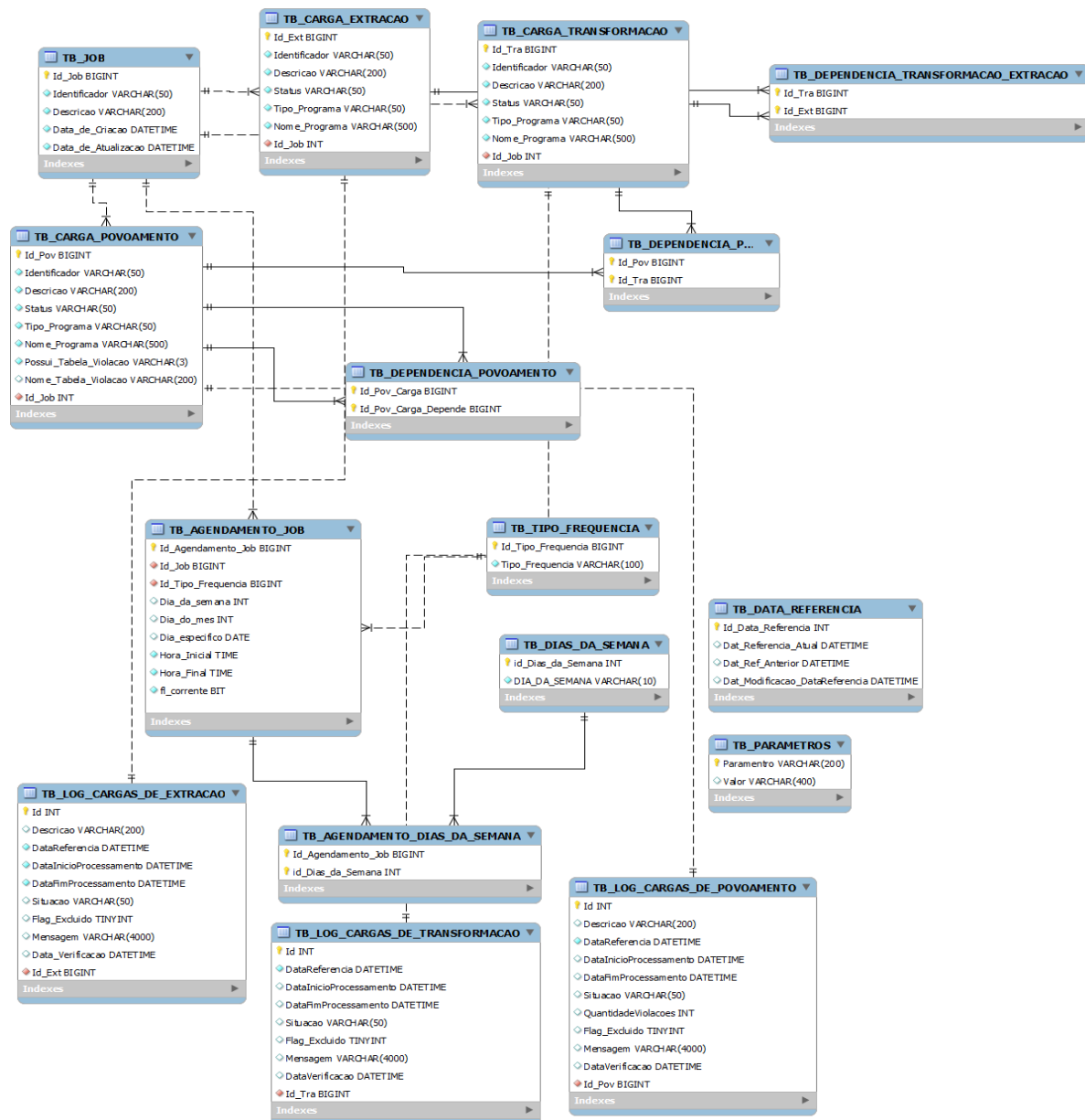


Figura 14: Esquema de Dados da Ferramenta

4. FERRAMENTA

A ferramenta tem como objetivo principal projetar e implementar um subsistema para agendamento, execução e monitoramento de processos ETL.

Nesta seção apresentaremos um estudo de caso com um esquema de dados criado para uma rede de cinemas, cujas tabelas fornecem informações sobre diversas transações de venda de ingressos e produtos. Esse esquema foi criado durante a disciplina Sistemas de Informação, para o projeto de Data Warehouse. Faremos uso do esquema do ambiente operacional, *CINEMA_OLTP*, para demonstrar o funcionamento da ferramenta.

Menu Principal

A Figura 15 apresenta o menu principal da ferramenta.

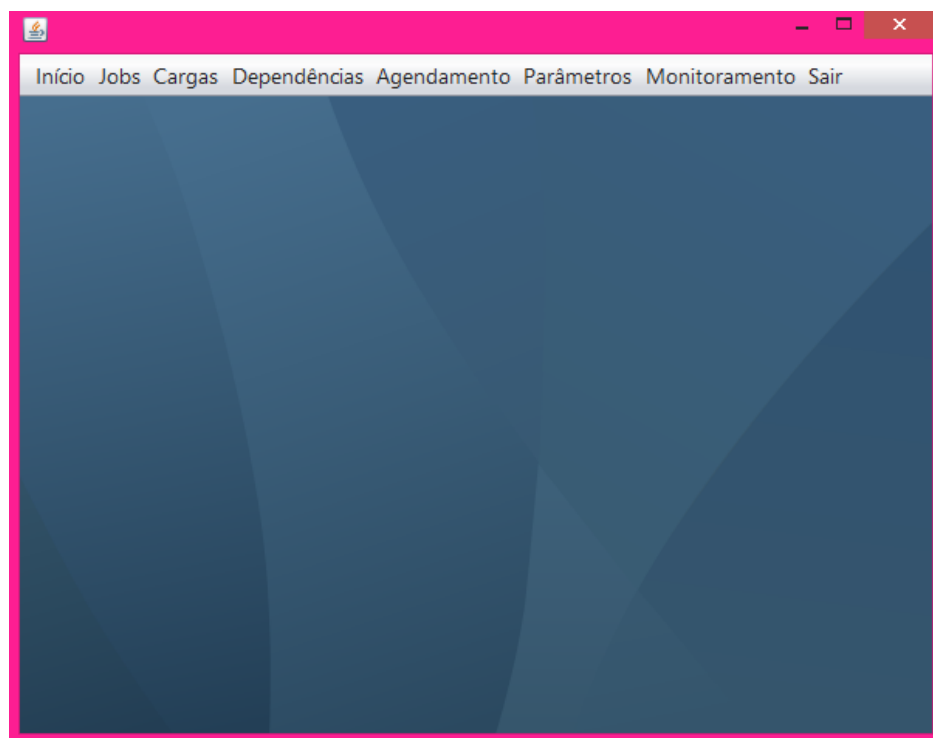


Figura 15: Tela do Menu Principal da Ferramenta

Criar Jobs

Na tela da figura 16, o usuário cadastra um novo Job. São inseridas informações como o identificador do Job, que seria uma pequena descrição. A descrição que vai detalhar qual a sua funcionalidade. A data de criação e a data de atualização são inseridas automaticamente pelo sistema. A data de criação é inserida quando o usuário cria o Job, e a data de atualização quando o Job for atualizado novamente. O usuário clica em *Cadastrar* para finalizar a criação, *Limpar* para refazer o cadastro ou *Fechar* para abrir encerrar o processamento.

Nesse caso, foi criado o *Job Dw Cinema*, que tem a função de realizar todo o povoamento da base Cinema usado como exemplo. O Job cadastrado tem a função de extrair os dados do ambiente OLPT, transformar e finalmente atualizar os dados finais no DW Cinema.

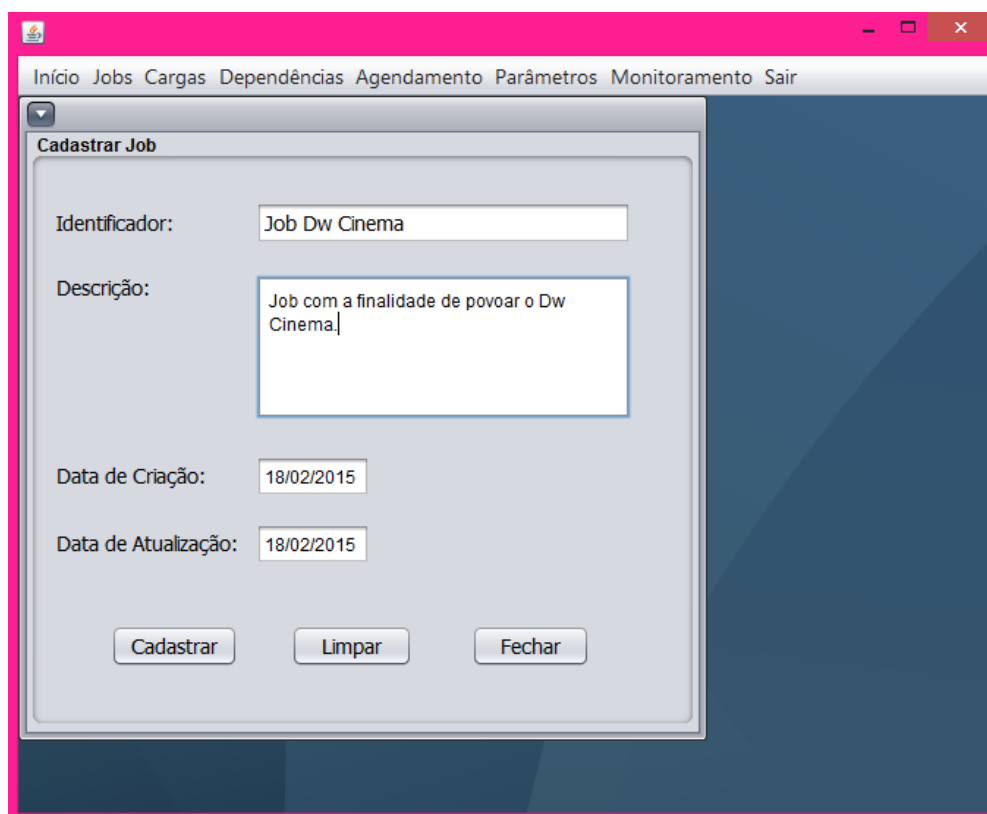


Figura 16: Tela de Cadastro do Job

Editar Jobs

Essa tela apresenta uma opção para o usuário selecionar qual o Job cadastrado ele deseja editar, representado pela figura 17. Ao clicar em *Editar* uma nova tela é

aberta com todos os dados referente ao Job escolhido. A data de atualização é alterada automaticamente para a data na qual o usuário está realizando a modificação e a data de criação permanece a mesma, como mostra a figura 18.

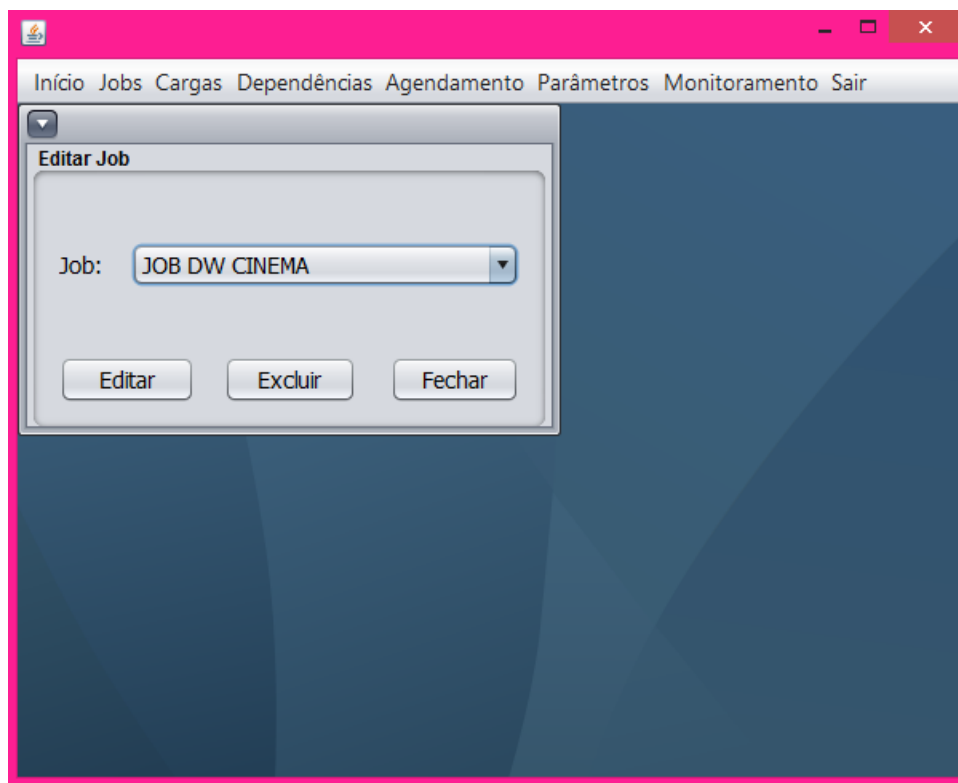


Figura 17: Tela de Editar Job da Ferramenta

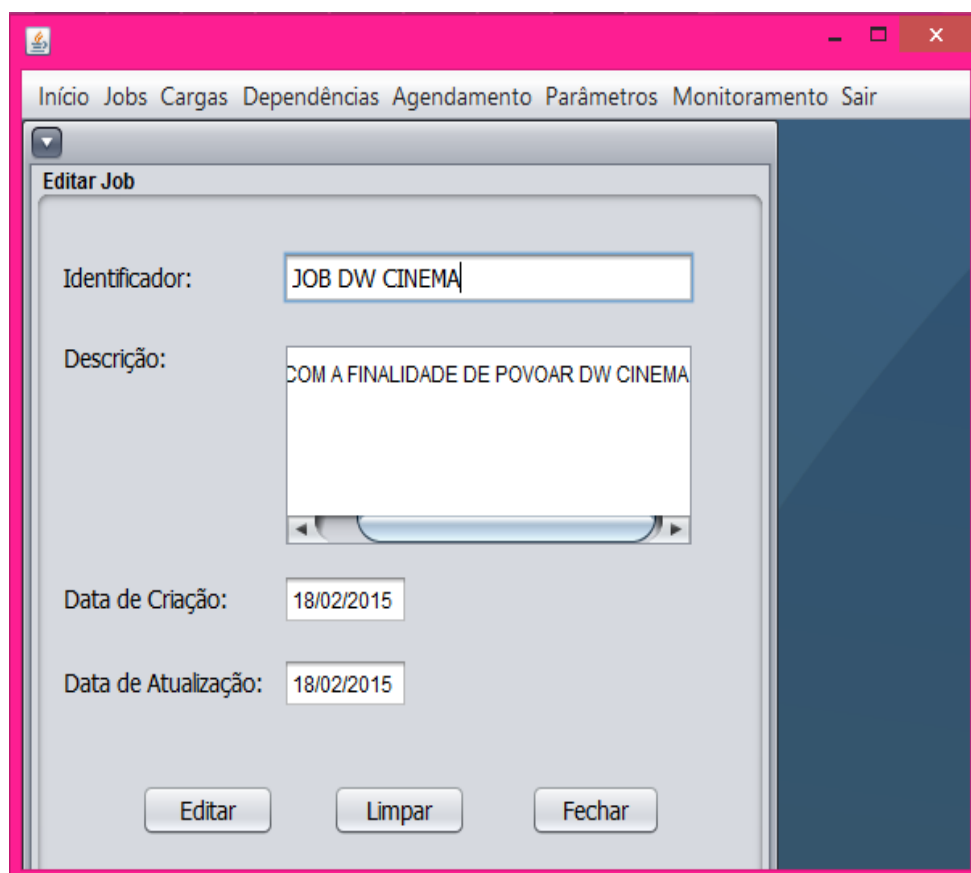


Figura 18: Tela de Editar Job da Ferramenta

Cadastrar Carga de Extração

Na seção das cargas ETL, a ferramenta apresenta funcionalidades para cadastro e edição das cargas de Extração, Transformação e Povoamento. A figura 19 representa a tela de cadastro da carga de Extração. Nessa tela o usuário seleciona o Job cadastrado anteriormente, nesse caso o *Job Dw Cinema*. Fornece uma breve descrição da carga no campo Identificador e uma descrição mais detalhada no campo Descrição. O usuário determina se a carga será agendada ou não através do status ATIVA ou INATIVA.

No caso abaixo foi cadastrada a carga de Extração *SP_EXTRAÇÃO_FILME*, como um procedimento armazenado quem tem a função de extrair os dados do ambiente *Cinema_OLTP* e migrar pra a Área de Staging.

Job: JOB DW CINEMA

Identificador: Extrair dados da tabela Filmes

Descrição: Essa carga tem a função de extrair todos os dados encontrados na tabela Filme do ambiente operacional CINEMA_OLPT.

Status: ATIVA

Tipo do Programa: Procedimento Armazenado

Nome do Programa: SP_EXTRAÇÃO_FILME

Cadastrar Limpar Fechar

Figura 19: Tela de Cadastro da Carga de Extração da Ferramenta

Editar Carga de Extração

Na tela, representada pela figura 20, o usuário seleciona qual o Job que está relacionado com a carga a ser editada, nesse caso o *Job Dw Cinema*. Após isso seleciona a carga de Extração *Extrair dados da tabela Filmes* e clica em *Editar*. A tela de edição apresenta todas as informações que foram registradas sobre a carga de extração, o usuário modifica as informações desejadas e clica em *Editar*. A figura 21 representa a tela de edição da carga de Extração.

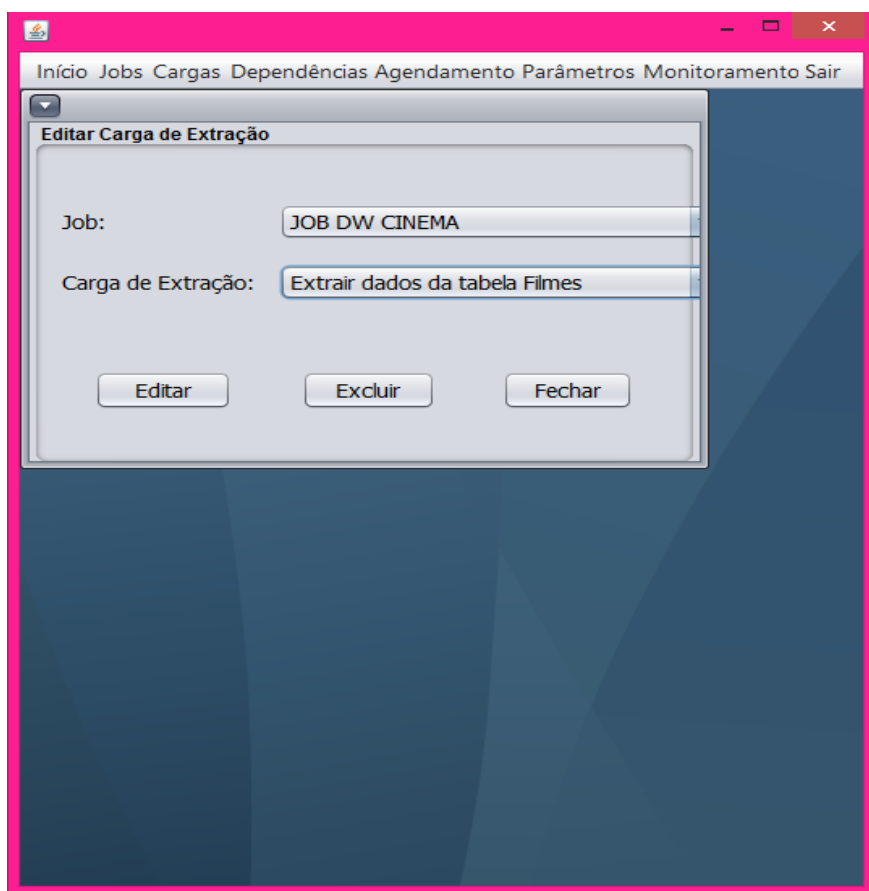


Figura 20: Tela de Edição da Carga de Extração da Ferramenta

Início Jobs Cargas Dependências Agendamento Parâmetros Monitoramento Sair

Editar Carga de Extração

Job: JOB DW CINEMA

Identificador: Extrair dados da tabela Filmes

Descrição: Essa carga tem a função de extrair todos os dados encontrados na tabela Filme do ambiente operacional CINEMA_OLPT.

Status: ATIVA

Tipo do Programa: Procedimento Armazenado

Nome do Programa: SP_EXTRAÇÃO_FILME

Editar Limpar Fechar

Figura 21: Tela de Edição da Carga Extração da Ferramenta

O mesmo processo ocorre com as cargas de Transformação e Povoamento. As figuras 22, 23, 24, 25, 26 e 27 representam as telas de cadastro e edição de ambas as cargas.

The image shows a software application window with a menu bar at the top containing the following items: Início, Jobs, Cargas, Dependências, Agendamento, Parâmetros, Monitoramento, and Sair. A dialog box titled 'Cadastrar Carga de Transformação' is open in the foreground. The dialog box contains the following fields and controls:

- Job:** A dropdown menu with 'JOB DW CINEMA' selected.
- Identificador:** A text input field containing 'Transformar dados da tabela Filmes'.
- Descrição:** A text area containing 'Realizar a limpeza e integração dos dados da tabela Filmes.'
- Status:** A dropdown menu with 'ATIVA' selected.
- Tipo do Programa:** A text input field containing 'Procedimento Armazenado'.
- Nome do Programa:** A text input field containing 'SP_TRANSFORMAÇÃO_FILMES'.

At the bottom of the dialog box, there are three buttons: 'Cadastrar', 'Limpar', and 'Fechar'.

Figura 22: Tela de Cadastro da carga de Transformação da Ferramenta

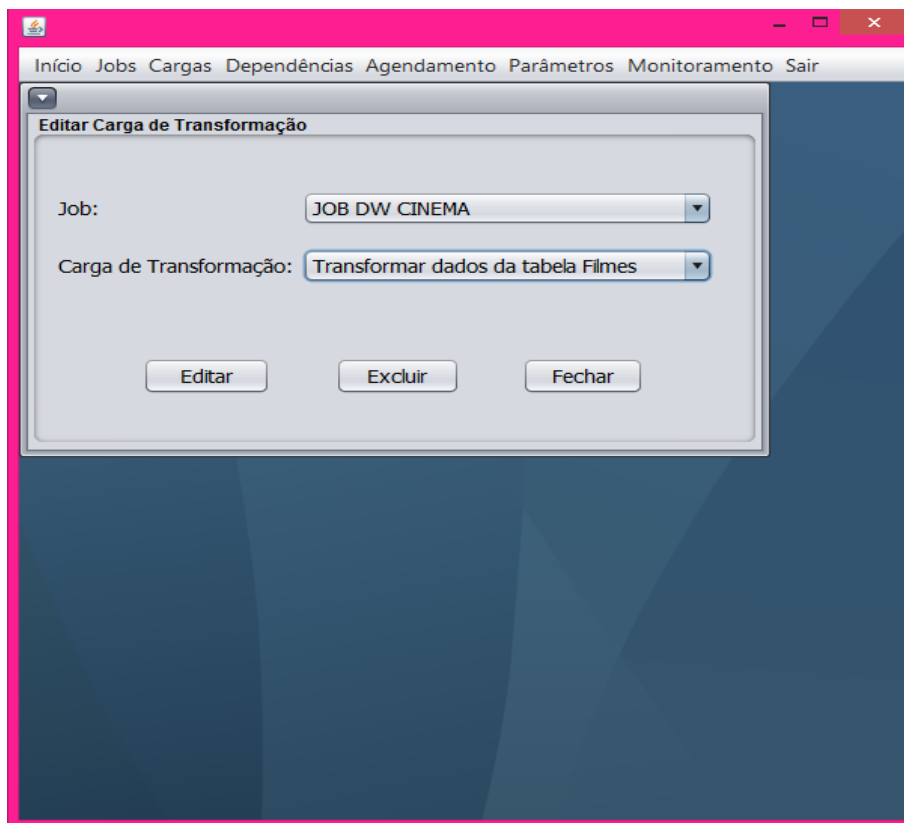


Figura 23: Tela de Edição da carga de Transformação da Ferramenta

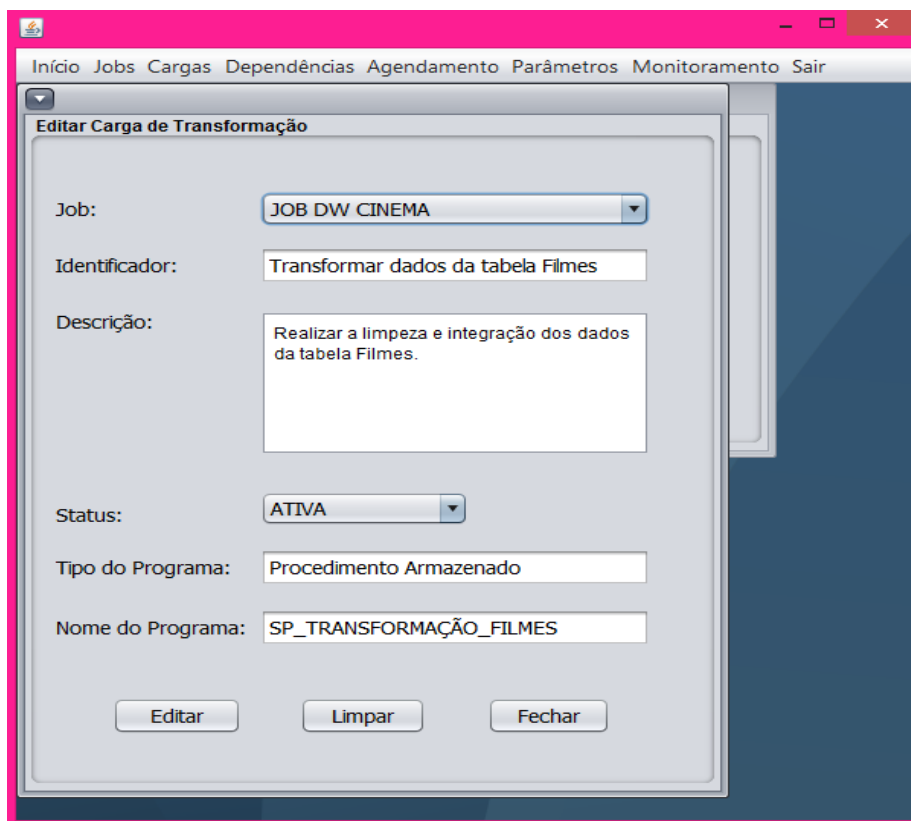


Figura 24: Tela de Edição da carga de Transformação da Ferramenta

Início Jobs Cargas Dependências Agendamento Parâmetros Monitoramento Sair

Cadastrar Carga de Povoamento

Job: JOB DW CINEMA

Identificador: Povoar a Dimensão Filmes

Descrição: Essa carga tem a função de povoar a dimensão Filmes no Dw Cinema com os dados da tabela auxiliar Filmes da área de Staging

Status: ATIVA

Tipo do Programa: Procedimento Armazenado

Nome do Programa: SP_CARGA_DIM_FILMES

☐ Possui Tabela de Violação

Cadastrar Limpar Fechar

Figura 25: Tela de Cadastro da carga de Povoamento da Ferramenta

Início Jobs Cargas Dependências Agendamento Parâmetros Monitoramento Sair

Editar Carga de Povoamento

Job: JOB DW CINEMA

Identificador: Povoar a Dimensão Filme

Descrição: Essa carga tem a função de povoar a dimensão Filmes no Dw Cinema com os dados da tabela auxiliar da área de Staging.

Status: ATIVA

Tipo do Programa: Procedimento Armazenado

Nome do Programa: SP_CARGA_DIM_FILME

☐ Possui Tabela de Violação

Editar Limpar Fechar

Figura 26: Tela de Edição da carga de Povoamento da Ferramenta

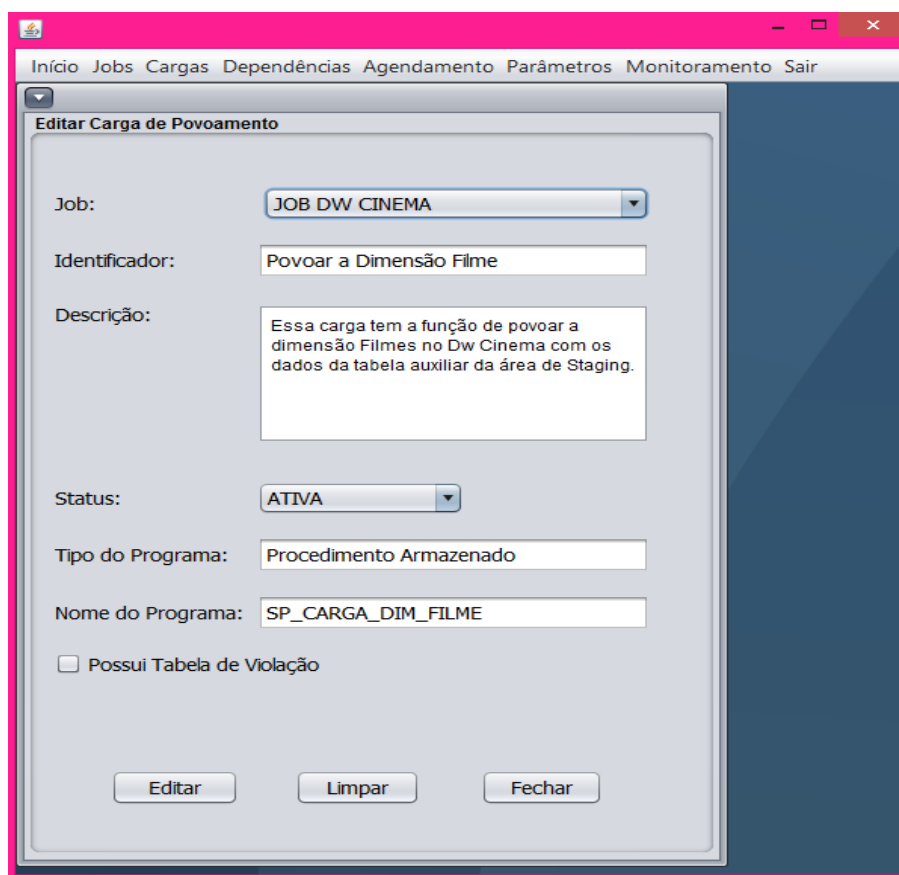


Figura 27: Tela de Edição da carga de Povoamento da Ferramenta

Vincular Cargas ETL

O sistema permite que o usuário possa determinar relações de dependências entre as cargas cadastradas. Para que o processo ETL funcione corretamente e adequadamente é necessário que seja estabelecida uma ordem no processamento de execução das cargas ETL. Tanto a carga de Transformação, bem como a carga de Povoamento necessitam que o processo de Extração ocorra primeiro, já que ela é responsável pela extração dos dados. Já a carga de Povoamento depende que após a extração dos dados, seja realizada a limpeza e integração dos dados, caracterizando o processo de Transformação.

Vincular cargas de Transformação e Extração

Essa tela permite que o usuário estabeleça a relação de dependência entre as cargas de Extração e Transformação. Logo no início da tela o usuário seleciona a carga de Transformação que deseja vincular com a carga de Extração. O sistema disponibiliza

a opção de vincular todas as cargas de Extração cadastradas ou somente as que forem selecionadas. A figura 28 representa a carga de Transformação *Transformar dados da Tabela Filme* vinculada com a carga de Extração *Extrair dados da tabela filmes*. Nesse caso a carga de Transformação só será executada após a carga de Extração.

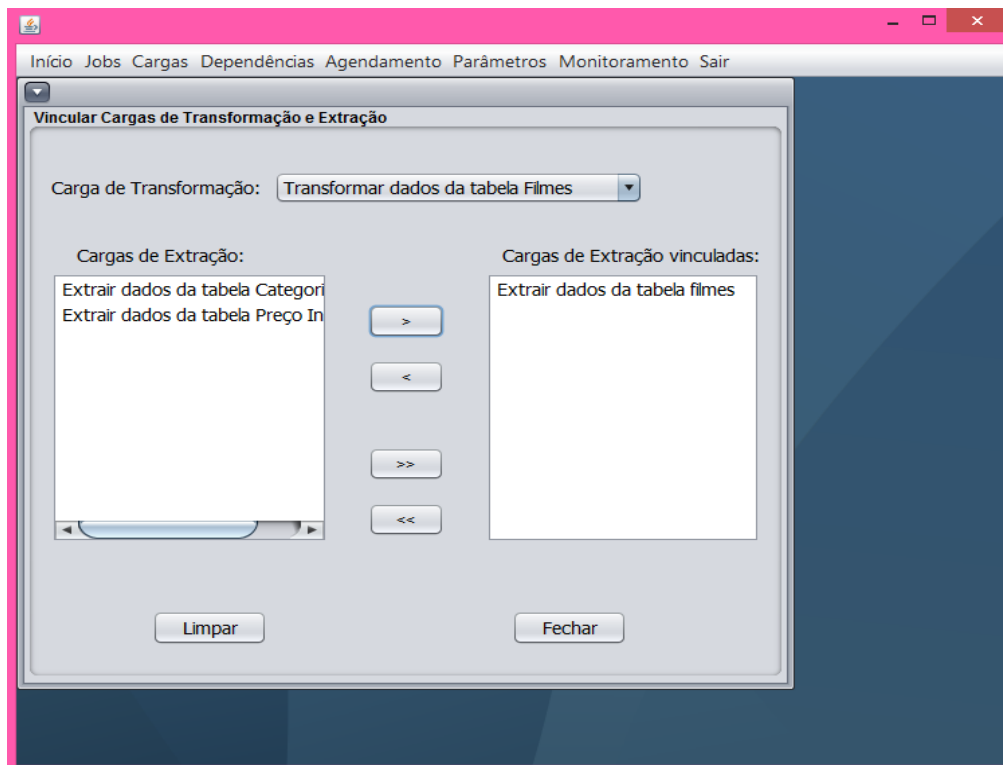


Figura 28: Tela de Vinculação entre a Carga de Transformação e Extração da Ferramenta

As figuras 29 e 30 representam o mesmo processo de vinculação com entre as cargas de Povoamento e Transformação e da carga de Povoamento.

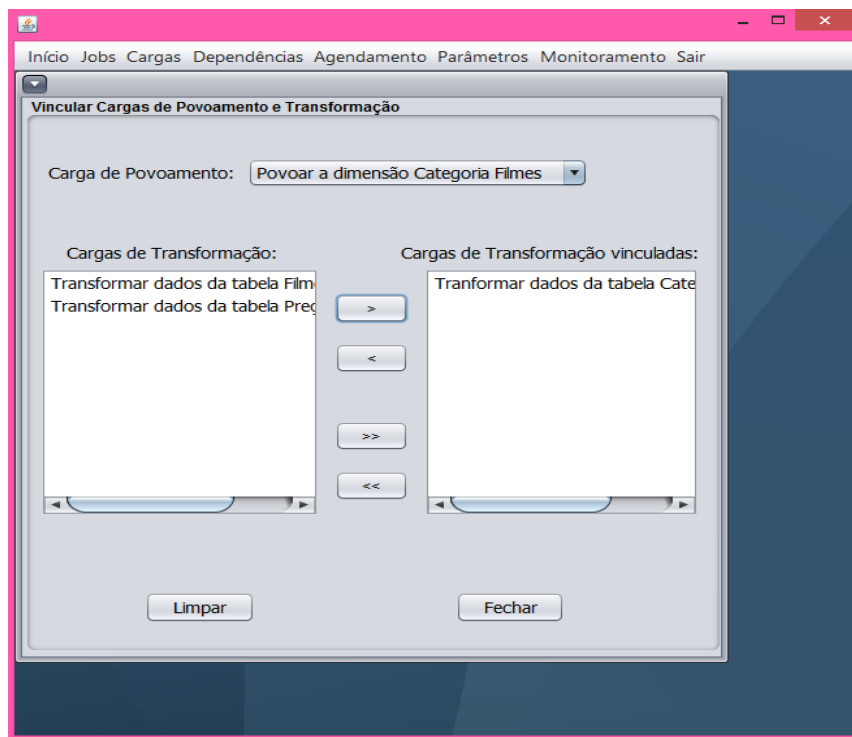


Figura 29: Tela de Vinculação entre a Carga de Povoamento e Transformação da Ferramenta

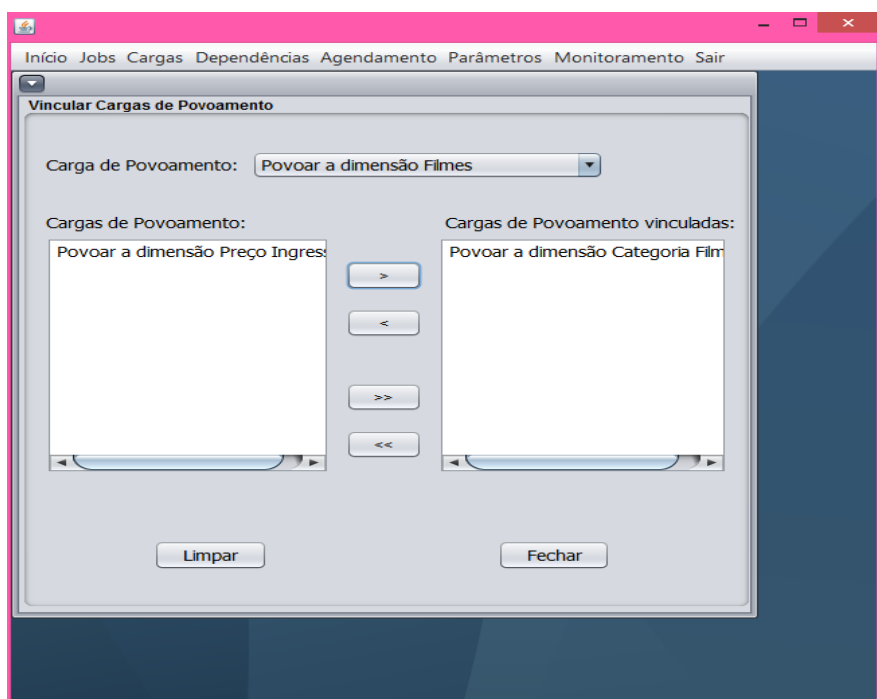


Figura 30: Tela de Vinculação entre as Cargas de Povoamento da Ferramenta

Cadastrar Agendamento

Na tela de agendamento representada na figura 31 o usuário seleciona o Job que deseja agendar. De acordo com o tipo de agendamento escolhido, os demais campos abaixo serão habilitados ou desabilitados. Caso o usuário deseje agendar o Job diariamente, os campos da Hora Inicial e Hora Final serão habilitados a fim de permitir a programação. Na opção semanalmente, os campos com os dias da semana estão disponíveis ao usuário. Outras opções escolhidas podem ser também ou um dia do mês ou data específica.

No estudo de caso abaixo o Job *Job Dw Cinema*, foi agendado para ser executado semanalmente, nos dias de segunda-feira e sexta-feira, das 08:00 às 10:00 da manhã.

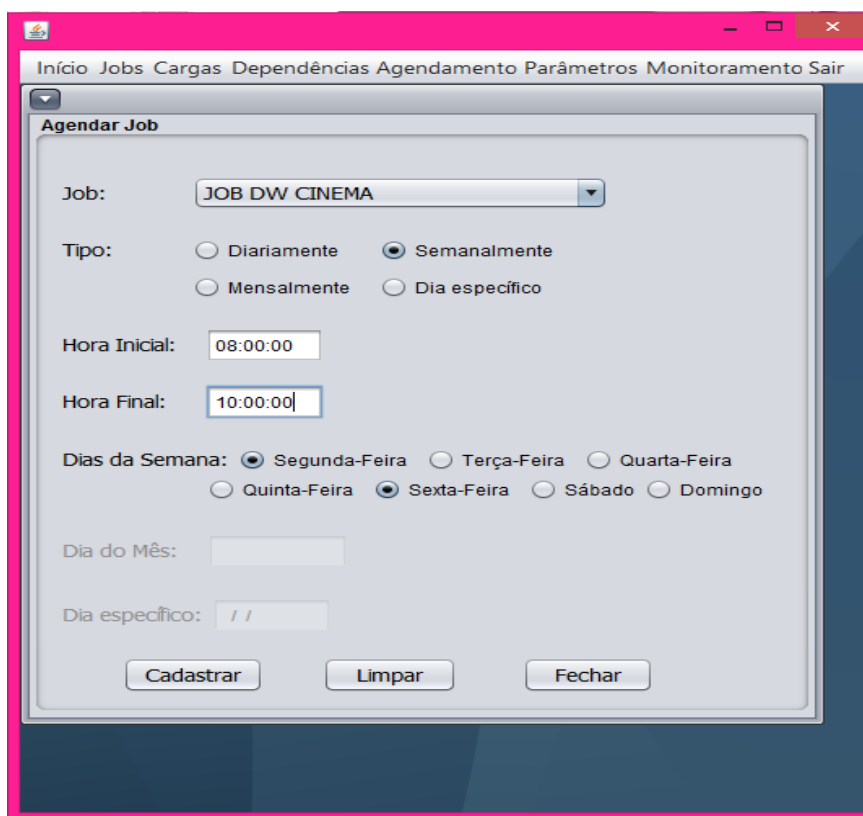


Figura 31: Tela de Agendamento do Job da Ferramenta

Editar Agendamento

Na tela 32 o usuário escolhe qual Job deseja editar o agendamento. Uma nova tela se abre, representada pela figura 33 com todos os dados referente ao agendamento do Job. O usuário altera os dados do agendamento e atualiza as novas informações, podendo também excluir o agendamento.

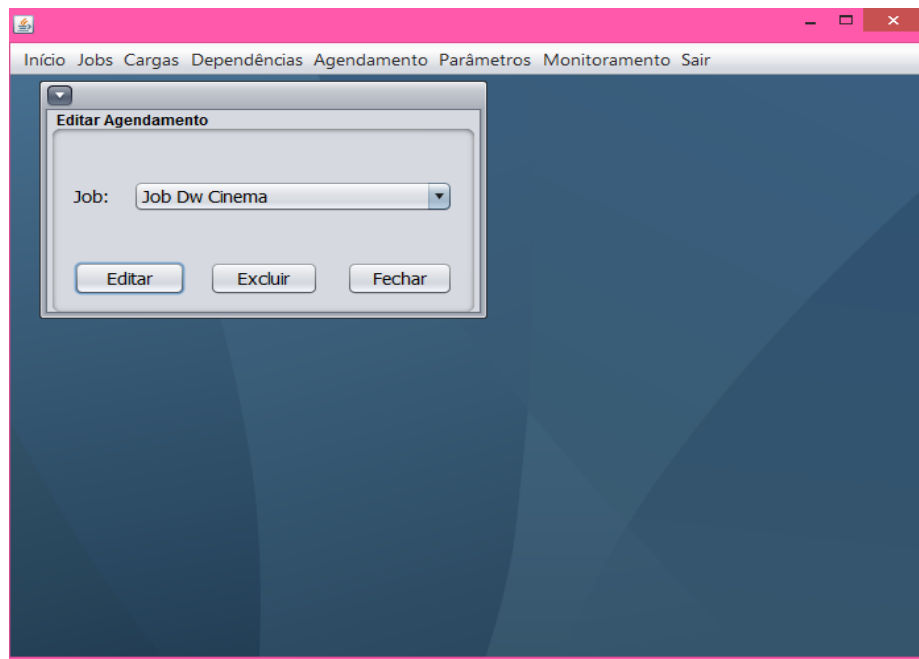


Figura 32: Tela de Edição do Agendamento do Job da Ferramenta

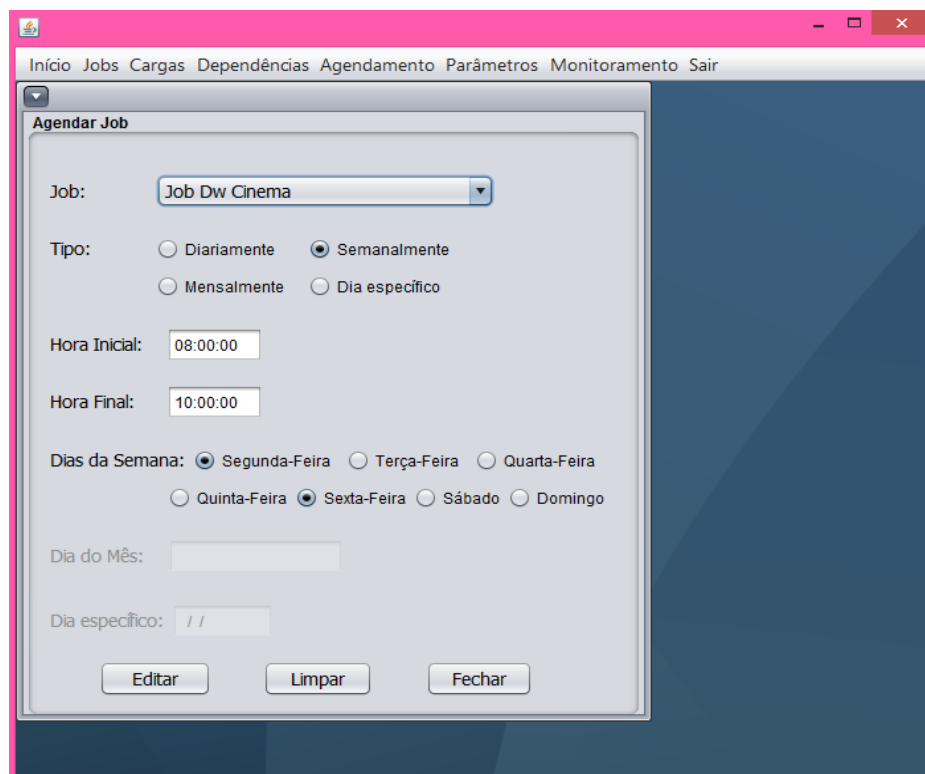


Figura 33: Tela de Edição do Agendamento do Job da Ferramenta

Cadastrar Parâmetro

A figura 34 representa a tela para cadastrar parâmetro. Nessa seção o usuário tem a possibilidade de cadastrar valores que podem ser utilizados pelo sistema.

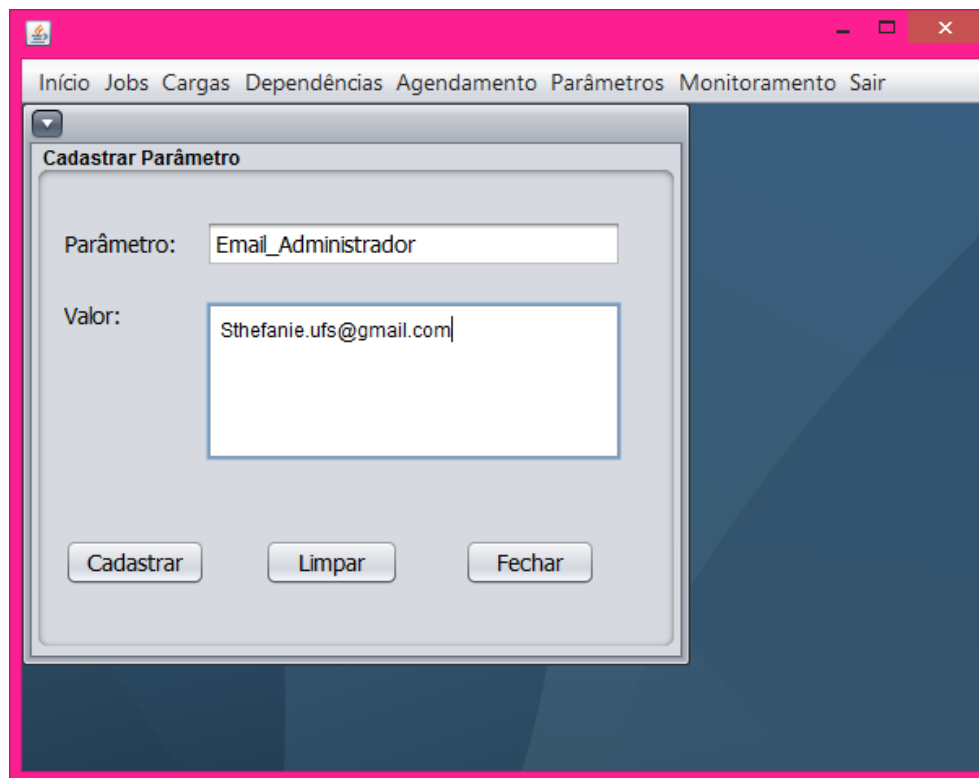


Figura 34: Tela de Cadastro dos Parâmetros da Ferramenta

Editar Parâmetro

Nesta seção o usuário seleciona na primeira tela qual o parâmetro que deseja alterar, representada pela figura 35. Após isso, a tela da figura 36, é aberta e o usuário edita as opções que desejar.

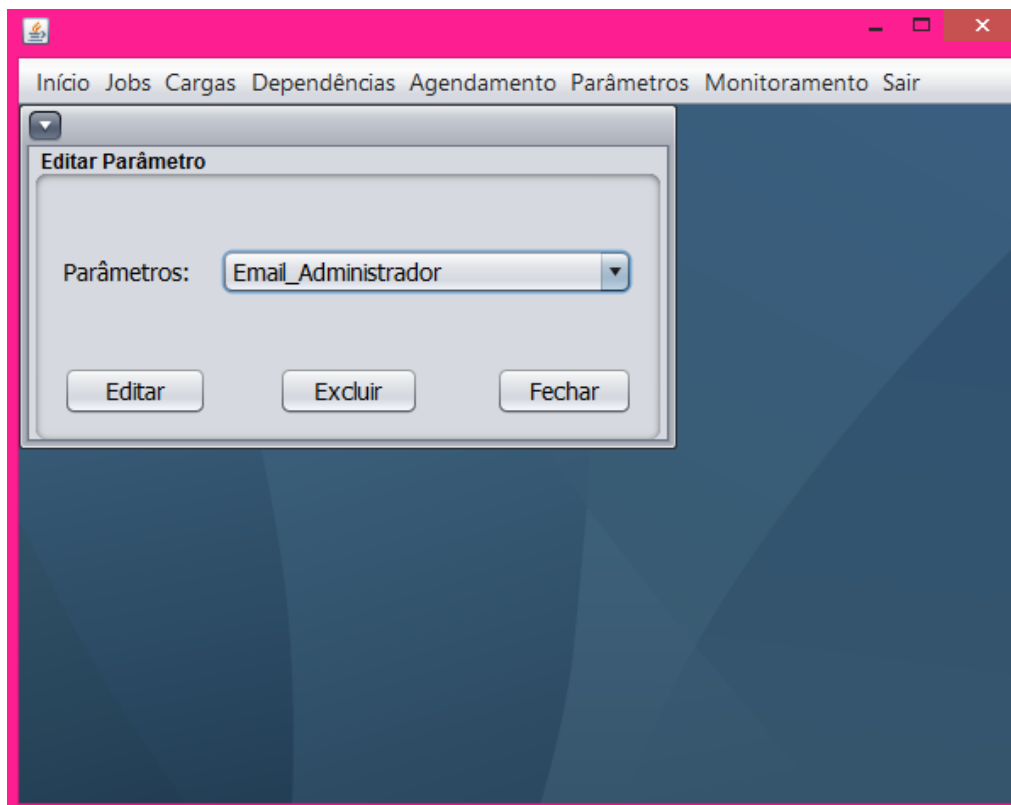


Figura 35: Tela de Edição dos Parâmetros da Ferramenta

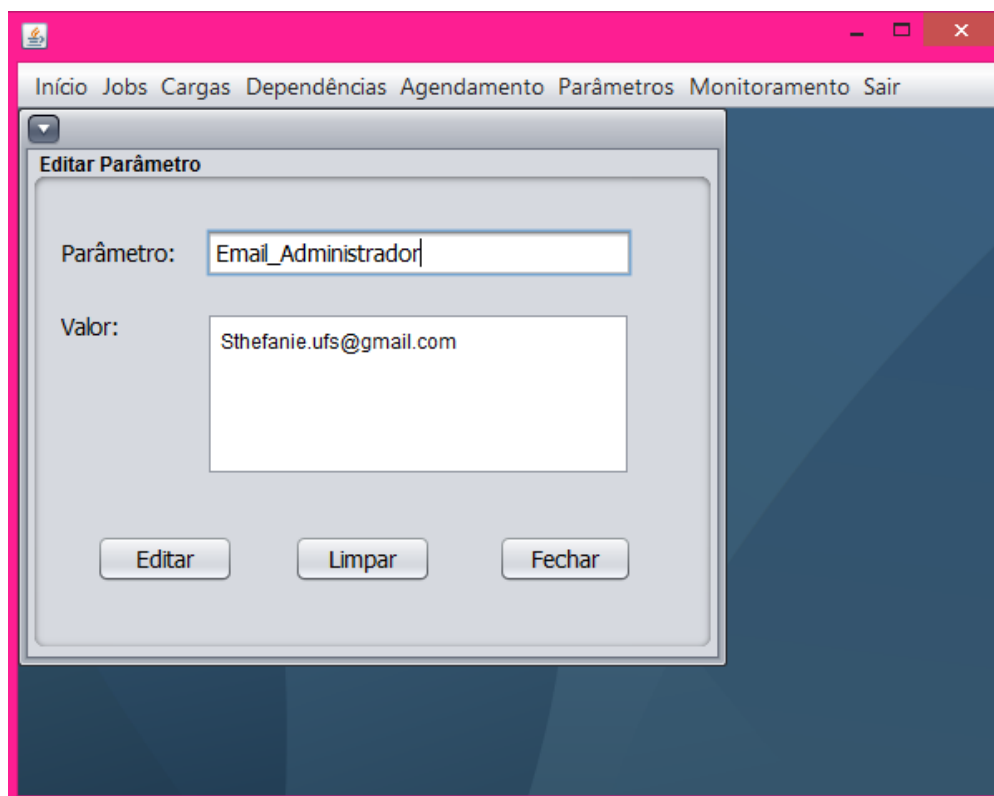


Figura 36: Tela de Edição dos Parâmetros da Ferramenta

Monitorar Cargas de Extração

Nessa tela, representada pela figura 37 o sistema permite listar todo o conteúdo referente às cargas de Extração que foram agendadas ou executadas pelo sistema. O usuário seleciona a data de referência a qual deseja listar, e as informações como por exemplo, descrição, mensagem, situação, data do início do processamento e do fim do processamento serão apresentadas na tela.

As figuras 38 e 39 representam o mesmo processo em relação às cargas de Transformação e Povoamento.

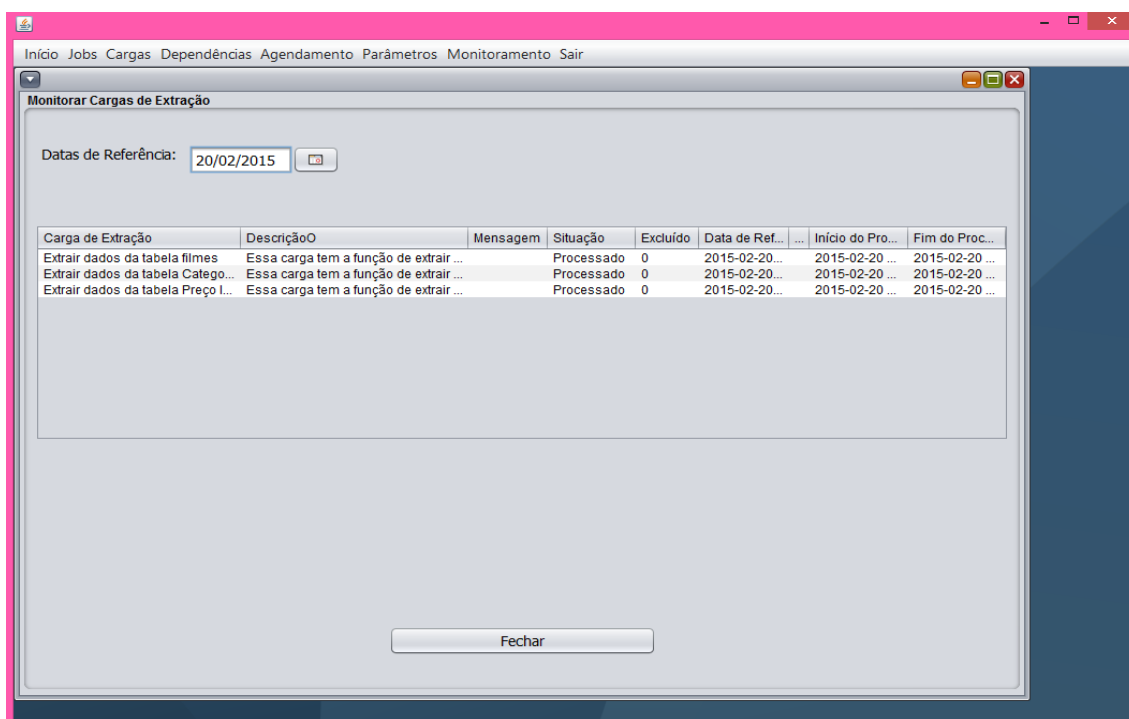


Figura 37: Tela de Monitoramento da Carga de Extração da Ferramenta

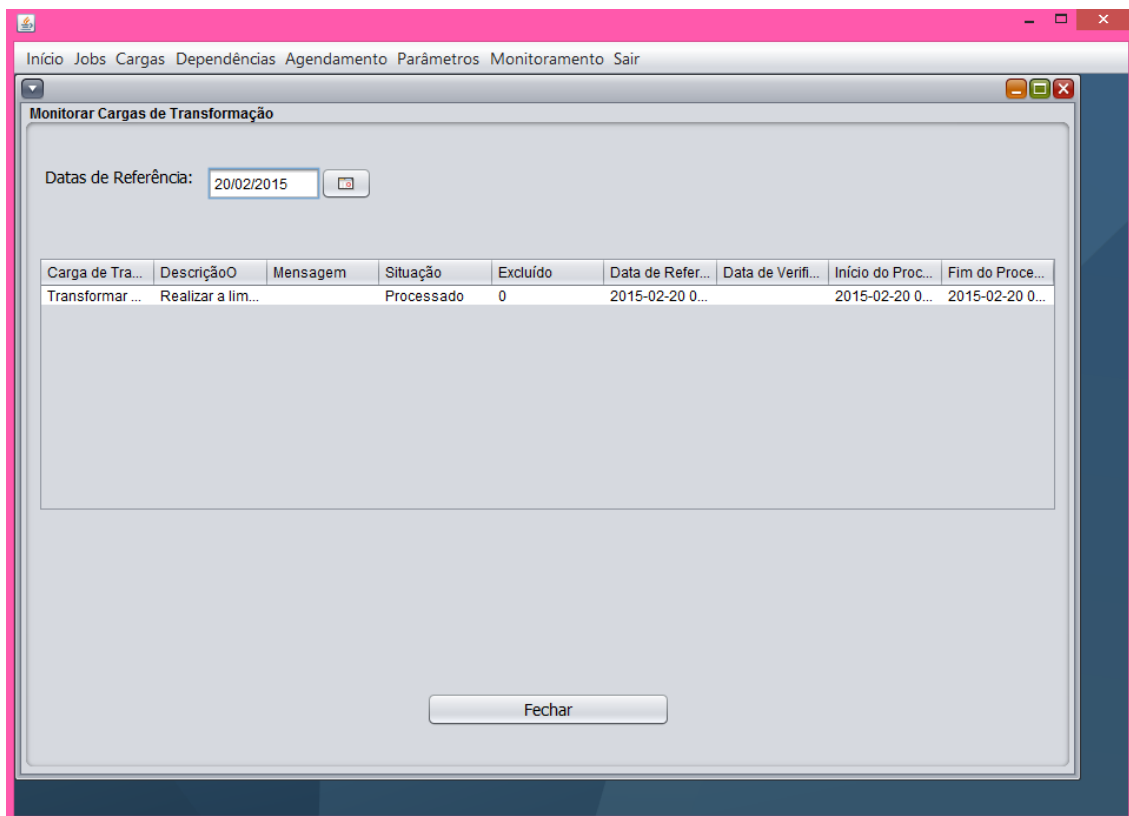


Figura 38: Tela de Monitoramento da Carga de Transformação da Ferramenta

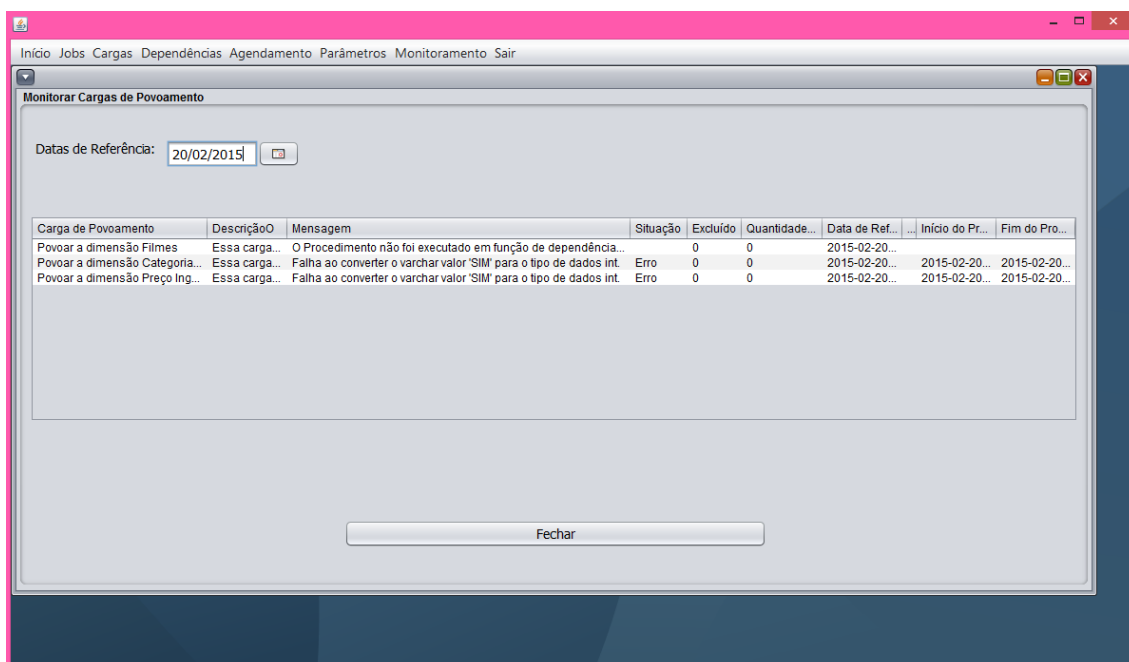


Figura 39: Tela de Monitoramento da Carga de Povoamento da Ferramenta

4.1 Procedimentos

O principal objetivo dessa ferramenta é o processo de agendamento das cargas ETL. Além do cadastro dos Jobs e das cargas de Extração, Transformação e Povoamento, o sistema possibilita o agendamento da execução de todas as cargas para os Jobs que tenham agendamento. Para tal foram criados três procedimentos responsáveis por realizar todo esse processo.

O primeiro procedimento serve para atualizar a data de referência. A data de referência nesse caso é a data para a qual execução das cargas será agendada. Para esse procedimento foi criada a tabela TB_DATA_REFERENCIA que tem o objetivo principal de identificar essa data. Nessa tabela são armazenados três campos principais, a data de referência atual, a data de referência anterior e a data de modificação. A data de referência atual é alterada diariamente e significa que toda carga será agendada e executada nessa data. A data de referência anterior é apenas um histórico e a data de modificação é apenas para saber quando a data atual foi atualizada.

O segundo procedimento é para agendar as cargas ETL. Nesse caso, as cargas serão agendadas em função da data de referência. O principal objetivo desse procedimento é agendar todas as cargas (extração, transformação e povoamento) para os Jobs que tenham agendamento para essa data. Ou seja, todos os Jobs que rodem diariamente, ou semanalmente (no caso da data de referência coincidir de ser durante essa semana) ou mensalmente caso a data de referência ocorra durante esse mês. Ou ainda no caso de dia específico. Esse procedimento ainda verifica a hora de início e fim. As cargas desses Jobs só poderão ser agendadas nesse intervalo de horário.

O terceiro procedimento é responsável por executar as cargas ETL que foram agendadas. O objetivo desse procedimento é recuperar as cargas agendadas nas tabelas de LOG e executar os procedimentos associados referente a cada carga. No caso de existir dependências entre as cargas, as cargas dependentes só poderão ser executadas após as cargas a qual ele depende serem concluídas.

5. CONCLUSÃO

O tema ETL (*Extract, Transform & Load*) não deve ser abordado apenas como um subprocesso na construção de um DW. Ele é muito mais que isso. Se uma organização possui uma área de TI com sistemas transacionais cujas informações necessitem ser migradas e analisadas, o uso do ETL se torna indispensável. Com o crescente volume de informações que as empresas têm que administrar nos dias atuais torna-se imprescindível a utilização de ferramentas ETL.

O principal desafio encontrado pelas grandes empresas ao adquirir uma ferramenta de integração de dados se dá por vários motivos. Um deles é o alto valor de suas licenças que dependendo da ferramenta pode chegar a 40% do custo de todo um projeto. Outro empecilho encontrado pelas empresas é a falta de profissionais capacitados no manuseio dessas ferramentas, que muitas das vezes podem se tornar bastante complexas. Nesse trabalho, propomos a construção de um *Scheduler* ETL. Trata-se de uma opção personalizada para projetar e implementar um subsistema para agendamento, execução e monitoramento de processos ETL. Dessa forma garantindo que os processos sejam executados no tempo certo, de acordo as relações e dependências existentes.

A principal função do *Scheduler* ETL é garantir que as atividades de agendamento, execução e monitoramento dos processos ETL sejam concretizadas no tempo certo, sempre respeitando o acordo entre as relações e dependências existentes. A presente ferramenta, proposta neste trabalho, difere das propostas anteriores por apresentar uma abordagem mais simples dos processos de agendamento, execução e monitoramento dos processos ETL. Buscando favorecer projetos de Ambientes de Suporte à Decisão que optem por não adquirir ferramentas ETL.

Não é pretensão deste trabalho julgar ou apresentar a melhor ferramenta de ETL. A principal contribuição desta pesquisa foi apresentar um modelo de ferramenta ETL customizado que possibilitasse de certa forma atender aos requisitos necessários para um determinado projeto. Seu principal objetivo é ser uma ferramenta de manuseio simples, munidos com metadados necessários à execução dos processos ETL, possibilitando o agendamento e monitoramento de todo os processos que alimentam um *Data Warehouse*.

O presente trabalho pode servir de base para trabalhos futuros. Como extensões desse trabalho, possibilitar a execução de diferentes tipos de programas, criar procedimentos ou programas de agendamento e execução que pudessem ser transportados para outras plataformas, testar a ferramenta em ambientes de laboratório através de experimentos.

REFERÊNCIAS

- ADAMSON, C. Mastering Data Warehouse Aggregates: Solutions for Star Schema
- ALMEIDA, ALEXANDRE MARQUES DE. Proposição de Indicadores Para avaliação Técnica de Projetos de Data Warehouse: Um estudo de caso no Data Warehouse da Plataforma Lattes. Universidade Federal de Santa Catarina. 2006.
- COLAÇO JÚNIOR, M. Projetando sistemas de apoio à decisão baseados em data Dimensional Modeling. 2. ed. John Wiley and Sons, Inc., 2002.
- GONÇALVES, MARCIO. Extração de Dados pra Data Warehouse. Rio de Janeiro: Axcel Books, 2003.
- IMHOFF, C.; GALEMMO, N.; GEIGER. J. G. Mastering Data Warehouse Design: Relational and Dimensional Techniques. Indianapolis: Wiley Publish, Inc., 2003. Indiana: Wiley Publishing Inc., 2005.
- INMON, W. H. Building the data warehouse, Fourth Edition. 4. ed. Indianapolis,
- INMON, W. H. Como Construir o Data Warehouse. Campus, Rio de Janeiro, 1997.
- KABIRI, AHMED; CHIADMI, DALILA. Survey on ETL Processes. Journal of Theoretical and Applied Information Technology., 2013.
- KIMBALL, R. Data Warehouse Toolkit. Tradução Mônica Rosemberg; Revisão Técnica Ronald Stevis Cassiolato. São Paulo: Makron Books, 1998.
- KIMBALL, R. The Data Warehouse ETL Toolkit. 1 ed. Wiley India (P) Ltd., 2004.
- KIMBALL, R; ROSS, M.; THORNTHWAITE, W. The data warehouse lifecycle toolkit. 2. ed. Indianapolis, Indiana: Wiley Publishing Inc., 2008.
- KIMBALL, R; ROSS, M. The data warehouse toolkit: The complete Guide to Performance. Indianapolis: Wiley Publish, Inc., 2006. warehouse. Rio de Janeiro: Axcel Books, 2004.

