Manual de arquitetura do Extrator de Dados

Sumário

Histórico de Alterações 3
Introdução 4
1 Fases da solução de integração 4
1.1 Fase 1 - Carga da Área de Staging 5
1.1.1 Fase 1a – Extração de Dados 6
1.1.2 Fase 1b – Transformação de Dados 6
1.2 Fase 2 - Carga das Tabelas Primárias do Sistema e-Gestão 6
1.3 Fase 3 - Geração das Remessas 7
1.4 Dependências entre as fases do processo 7
2 Arquitetura 7
2.1 Módulo de Carga da Área de Staging 8
2.2 Módulo de Cargas das tabelas primárias do Sistema e-Gestão 8
2.3 Módulo de Indicadores 8
2.4 Módulo de Geração das Remessas 8
3 Desenvolvimento da solução 8

Histórico de Alterações

Versão	Descrição	Data	Responsável
1.0	Versão Inicial	20/10/2015	Livia Matos/Gustavo Orair

Introdução

Este documento apresenta a arquitetura do Módulo Extrator de Dados do Sistema de Gerenciamento de Informações Administrativas e Judiciárias da Justiça do Trabalho (e-Gestão). Este sistema provê a solução de integração de dados entre os sistemas PJe e e-Gestão.

A integração de dados entre os sistemas PJe e e-Gestão é realizada por três fases principais:

1. Carga da Área de Staging
    a. Extração de dados
    b. Transformação de dados
2. Carga das tabelas primárias
3. Geração de remessas

Este documento apresenta os módulos que compõem a solução assim como as responsabilidades de cada módulo e as fases do processo de integração que cada módulo é utilizado.

1 Fases da solução de integração

Os dados que são pertinentes para a geração de remessas do e-Gestão são extraídos da base de dados do PJe (bases de dados transacionais) e são copiados em uma nova base de dados (chamada de Área de Staging). As informações constantes nesta base de dados são então lidas, transformadas e enriquecidas, com foco analítico, de acordo com as regras de negócio do sistema e-Gestão. Estas informações são então utilizadas para a posterior geração de remessas para o sistema e-Gestão.

No esquema abaixo é apresentada uma visão simplificada das fases do processo de geração das remessas do e-Gestão.

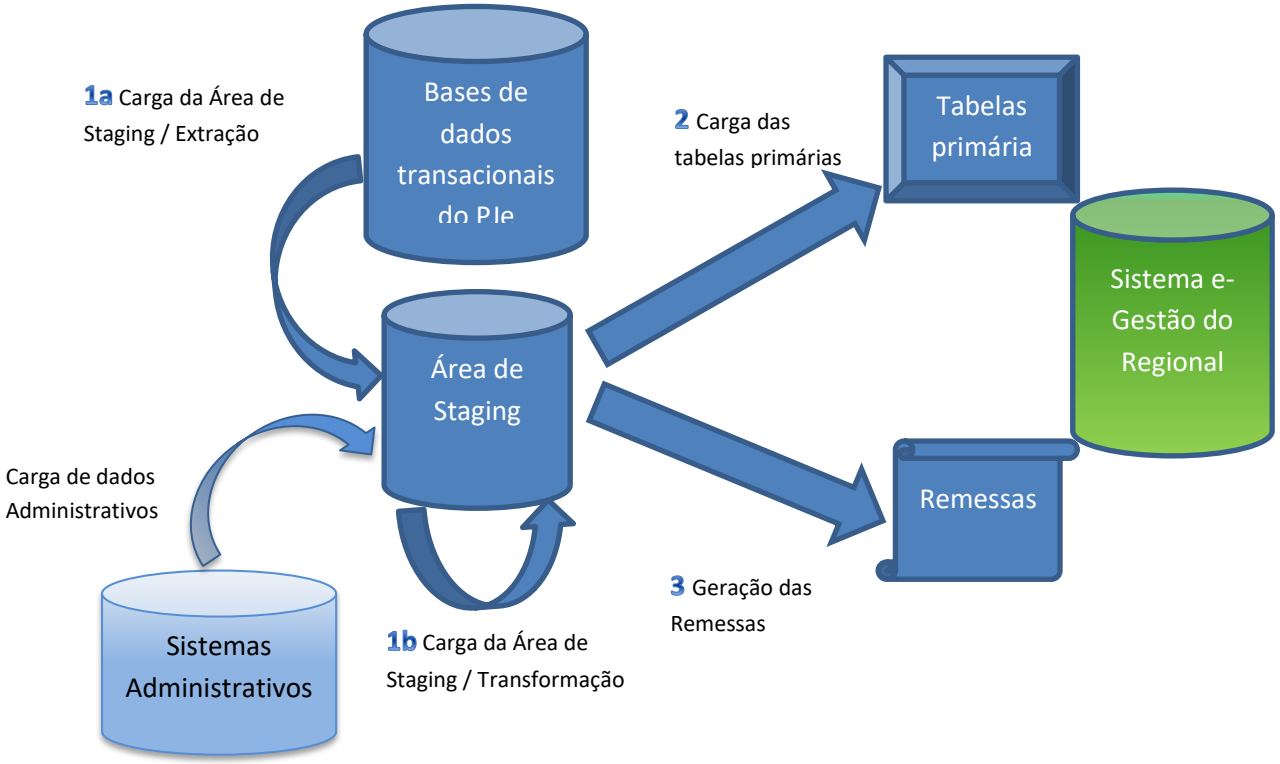


Figura 1 - Fases do processo de carga da área de staging, carga das tabelas primárias e geração de remessas

1.1 Fase 1 - Carga da Área de Staging

Nesta fase do processo, as informações das bases de dados (transacionais) do Sistema PJe são extraídas, transformadas e enriquecidas, com foco analítico, e posteriormente armazenadas na Área de Staging.

O processo de integração de dados da solução possui características de ELT em que os dados são extraídos das bases de dados do sistema PJe antes de iniciar a transformação e enriquecimento das informações – em que se manipulam dados exclusivamente da área de Staging. Por isto, a fase de carga da Área de Staging está dividida em duas subfases principais:

1. Extração de dados
2. Transformação de dados

Durante a carga da Área de Staging também pode haver necessidade de se realizar a carga de outros dados de outros sistemas. Como por exemplo, a carga dos dados dos afastamentos dos

magistrados do sistema RH do órgão. Essas informações são consultadas e utilizadas em fases posteriores à carga, como a geração das remessas.

1.1.1 Fase 1a – Extração de Dados

Esta subfase realiza a extração dos dados do sistema PJe pertinentes às análises do sistema e-Gestão. Esta é a única fase em que as bases de dados transacionais (do sistema PJe) são acessadas.

Visando otimizar o tempo de execução desta fase, diminuir as consultas ao servidor de banco de dados do PJe e a quantidade de dados trafegados, as seguintes técnicas de transferência de dados podem ser utilizadas:

- Carga Total
  Neste tipo de carga os dados são transferidos na sua totalidade. Ou seja, as tabelas da base de dados destino serão apagadas ("truncadas") e recarregadas totalmente.
- Carga Incremental
  Neste tipo de carga, os dados de destino são incrementados apenas com a adição dos novos registros oriundos da base de dados fonte. Ou seja, não é necessário transferir todos os dados novamente, apenas os novos.

1.1.2 Fase 1b – Transformação de Dados

Esta subfase concentra as transformações necessárias para uma modelagem analítica das informações. Assim, as informações já são armazenadas com base em um modelo analítico que permite simplificar as fases subsequentes do processo, assim tais fases podem realizar consultas de forma simplificada.

Nesta fase são realizadas várias ações sobre os dados, como a limpeza de dados – por exemplo, a alteração da data de autuação do processo por questões de inconsistências com a data da distribuição, a alteração das classes do processo caso o sistema PJe esteja configurado com uma classe do processo civil ao invés do processo trabalhista - , a vinculação de conceitos não existentes no modelo transacional, dentre outras várias ações necessárias para a apuração da estatística de forma consistente e acurada.

1.2 Fase 2 - Carga das Tabelas Primárias do Sistema e-Gestão

Nesta fase realiza-se a leitura de informações sobre os processos da Área de Staging tais como assuntos, indicadores dentre outras informações básicas e carregam-se tais informações no Sistema e-Gestão.

São preenchidas informações como os processos existentes em cada órgão estatística (EGT_PROCESSO), os assuntos dos processos (EGT_PROCESSO_ASSUNTO), indicadores dos processos (EGT_PROCESSO_INDICADOR) além de outras informações.

Note que tais informações básicas sobre os processos não possuem relação com as remessas. São tabelas "estáveis", pois independem das remessas, que compõem o modelo de dados do Sistema e-Gestão.

1.3 Fase 3 - Geração das Remessas

Nesta fase é realizada a extração dos indicadores do sistema e-Gestão e as informações obtidas são então transformadas no formato de remessas de acordo com o modelo de dados do e-Gestão.

Nesta fase são realizadas tarefas de mapeamento como, por exemplo, obter a classe no sistema e-Gestão de certa classe CNJ para o correto preenchimento das remessas. Por fim, as remessas geradas são carregadas no banco de dados do Sistema e-Gestão do Regional.

1.4 Dependências entre as fases do processo

É importante entender a dependência de dados entre as diferentes fases do processo.

Nota-se que a fase de carga da Área de Staging carrega todos os dados da Área de Staging e tais dados são consultados nas duas outras fases.

Assim existe uma dependência direta de dados entre as fases 2 e 3 à fase 1.

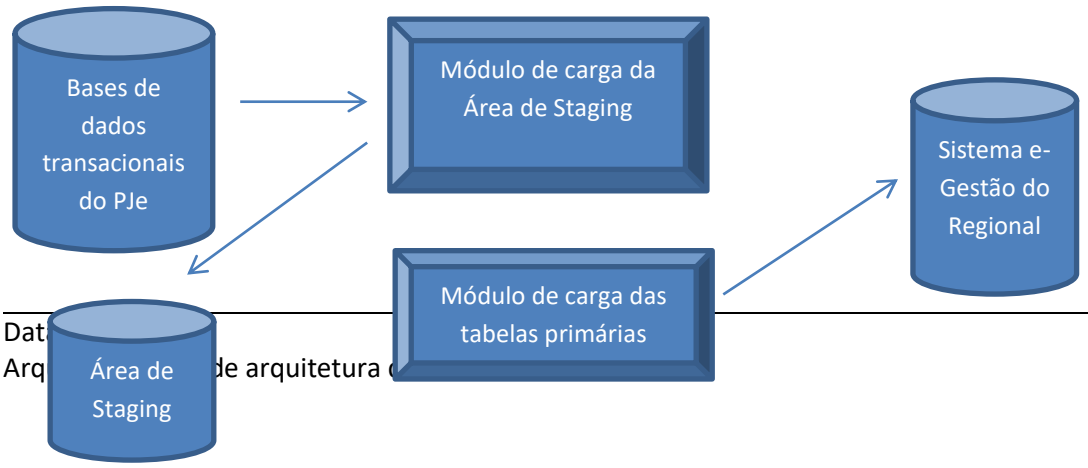
Além desta dependência direta, também existe dependência indireta de dados entre as fases 2 e 3. Isto acontece porque, conforme mostrado na **Figura 1 - Fases do processo de carga da área de staging, carga das tabelas primárias e geração de remessas**, a tabela EGT_PROCESSO do sistema e-Gestão é preenchida na fase de carga das tabelas primárias.

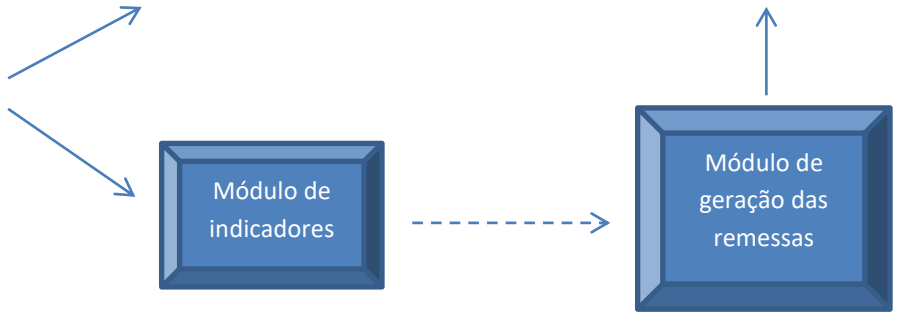
Assim, durante a fase de geração das remessas, são informados processos que já devem estar inseridos nesta tabela sob pena de problemas de restrições de integridade referencial no banco de dados do Sistema e-Gestão o que impede a correta execução da remessa.

2 Arquitetura

A arquitetura da solução se baseia nas fases existentes no processo de ETL.

A seguir apresentamos uma visão esquemática das bases de dados e dos módulos envolvidos na arquitetura:





2.1 Módulo de Carga da Área de Staging

Este módulo implementa as atividades/transformações de dados a serem executadas na fase de carga da Área de Staging.

2.2 Módulo de Cargas das tabelas primárias do Sistema e-Gestão

Este módulo é responsável por realizar a leitura dos processos, assunto, indicadores dentre outras informações básicas sobre os processos e carregar tais informações no Sistema e-Gestão.

2.3 Módulo de Indicadores

O módulo de indicadores implementa a lógica dos indicadores do e-Gestão. Ele é composto de uma série de SQLs parametrizados de consultas, uma para cada indicador do e-Gestão.

Este módulo foi desenvolvido de forma desacoplada do resto da solução o que permite que o mesmo seja utilizado por outros sistemas que tenham a necessidade de utilizar as consultas de forma mais customizada. Por exemplo, o SICOND – Sistema de Consulta a Dados Operacionais utiliza este módulo para realizar a consulta na Área de Staging para a exibição de Consultas com períodos diferenciados dos existentes nas remessas do e-Gestão.

2.4 Módulo de Geração das Remessas

Este módulo é o responsável pelo gerenciamento do processo de geração das remessas no sistema e-Gestão. Para isto, realiza a obtenção dos órgãos estatística (varas ou TRT) cujas remessas devem ser geradas, do período de referência da remessa, dos indicadores a serem executados. Ainda obtém as consultas dos indicadores (acesso ao módulo de Indicadores) e executa estas consultas com os parâmetros corretos de acordo com a remessa solicitada pelo usuário, transforma e carrega os indicadores na base de dados do e-Gestão.

3 Desenvolvimento da solução

A solução de integração de dados entre os sistemas PJe e e-Gestão foi desenvolvida utilizando uma ferramenta de ETL bem conhecida no mercado, o Pentaho Data Integrator (PDI).

As ferramentas de ETL objetivam a extração de dados de diversos sistemas, transformação desses dados conforme regras de negócios e por fim a carga dos dados geralmente em um Data Mart e/ou um Data Warehouse.

A utilização de uma ferramenta ETL no desenvolvimento da solução de integração de dados possibilita a construção e parametrização da solução de forma mais flexível. A estrutura da solução se baseia na utilização de repositórios PDIs específicos para cada módulo da solução (repositório para a carga da Área de Staging, carga das tabelas primárias e geração de remessas).

Estes repositórios são compostos por Jobs e Transformações ETL que podem ser executados a partir de uma interface gráfica (Spoon) ou podem ainda ser agendados em horários que melhor atendam às necessidades do Regional para execução em *batch (*Kitchen*)*.

As configurações de acesso às bases de dados destes repositórios são realizadas de forma altamente parametrizável. Assim, o Regional pode configurar as conexões (*Connections*) destes repositórios adequando-se facilmente a diferentes ambientes de acordo com a arquitetura física dos seus bancos de dados.