NOTA: Todas as instruções deste documento estão escritas utilizando o estilo “Commentary”. Você pode esconder e reexibir as instruções apenas clicando o botão de Parágrafo que aparece na barra de menu do Word. Quando o documento for finalizado você pode remover as instruções permanentemente utilizando a opção de “Delete Commentary”

Especificação Técnica de ETL

Versão 1.1

17/10/2012

Histórico de Revisões

|  |  |  |  |
| --- | --- | --- | --- |
| Data | Versão | Preparado por | Comentários |
| 04/10/2012 | 1.0 | Vanessa Felix | Versão Inicial do documento |
| 08/10/2012 | 1.0 | Leticia | Revisão e entrega do documento |
| 09/10/2012 | 1.0 | Marcelo Marques | Revisão do documento |
| 17/10/2012 | 1.1 | Vanessa | Ajuste do documento de acordo com a revisão do cliente |

Índice

[Histórico de Revisões 1](#_Toc338258124)

[1 Objetivo 2](#_Toc338258125)

[2 Escopo 3](#_Toc338258126)

[2.1 Descrição do Escopo 3](#_Toc338258127)

[2.1.1 Requisitos Funcionais 3](#_Toc338258128)

[2.1.2 Requisitos Não Funcionais 3](#_Toc338258129)

[2.2 Fora do Escopo 3](#_Toc338258130)

[2.3 Premissas 4](#_Toc338258131)

[3 Visão Lógica do Fluxo de Dados 4](#_Toc338258132)

[3.1 Fluxo Padrão 5](#_Toc338258133)

[3.1.1 Estruturas de Dados - m\_alocacao\_extensao 6](#_Toc338258134)

[4 Estratégias e Controles 8](#_Toc338258135)

[4.1 Controle e Tratamento de Exceções 8](#_Toc338258136)

[4.2 Expurgo de Dados 8](#_Toc338258137)

[4.3 Reinício e Recuperação 9](#_Toc338258138)

[4.4 Envio de Notificações 9](#_Toc338258139)

[5 Definição de Abreviaturas, Siglas e Acrônimos 11](#_Toc338258140)

1. Objetivo

O documento atual descreve a primeira etapa do processo padrão de coleta das informações dos sistemas de origem para carga no BI da Estácio. Neste processo as informações são extraídas de forma full ou incremental das tabelas dos sistemas de origem e inseridas na área de stage.

1. Escopo
   1. Descrição do Escopo

Este documento tem como objetivo estabelecer as definições padrões para desenvolvimento das interfaces de dados que realizaram extrações dos sistemas de origem e enviaram estes dados para as suas respectivas tabelas de Stage. Este processo será realizado para todas as tabelas, ou arquivos, envolvidas nos processos de ETL 1 da arquitetura de BI e na primeira etapa dos processos integração de dados em batch.

O processo de criação das tabelas de Stages utilizará, quando disponível, o script das tabelas do sistema de origem, modificado para conter os mesmos atributos e datatypes das tabelas de origem, porem sem as informações de foreign key, constraint e indexes. Alem dos respectivos campos de origem, serão acrescidos 4 novos campos a estas tabelas de Stages. São eles:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nome do campo | Tipo | Descrição do campo | Permite Nulos | Regra de Preenchimento |
| ID\_OCORRENCIA | VARCHAR2(20) | Identificador da execução do processo | Não | P\_ID\_OCORRENCIA |
| DATA\_REF | DATE | Data de referencia de carga do registro | Não | P\_DATA\_REF |
| STATUS\_PROC\_ETL2 | VARCHAR2(1) | Indica registros não processados pelo ETL 2 | Não | Valor fixo 'N' |
| ORIGEM\_INFORMACAO | VARCHAR2(20) | Indica a origem da informação | Não | Sigla do sistema de origem. Ex: SIA, Manual, CSV, ADP, SAP, ..c |

A relação contendo o nome de todas as tabelas ou arquivos de origem a serem utilizadas para a criação destes processos deverá ser fornecida por cada projeto no documento de solicitação de criação de processos de ETL1.

Os processos são incrementais por data de referência de processamento. Desta forma, apenas os registros alterados ou inseridos desde a última extração serão movimentados da origem para a área de stage. A forma padrão como esta extração será realizada, para cada um dos mapas, está descrita no item 3.1.

* + 1. Requisitos Funcionais

Este documento faz referencia a regra funcional RF34 do documento Requisitos Funcionais.xls

* + 1. Requisitos Não Funcionais

Este documento faz referência aos requisitos não funcionais RN02, RN03, RN04, RN06 e RN07 da planilha Requisitos Não Funcionais.xls.

* 1. Fora do Escopo

N/A

* 1. Premissas
* Apenas uma única tabela de stage pode ser criada para cada tabela ou arquivo de uma determinada origem.
* Execução com êxito do respectivo mapa de controle inicial (INICIA\_CONTROLE)
* Recebimento de parâmetros validos com as informações ID\_OCORRENCIA (identificador da execução) e DATA\_REF (data de referência)
* Para uma carga full, o parâmetro DATA\_REF deverá ser configurado com valor inferior ao menor valor de data existente no sistema de origem.
* Para carga incremental o parâmetro DATA\_REF deverá conter a data de referência do último processo executado com êxito.
* Os registros da origem a serem extraídos, tanto na carga full quanto na incremental, devem apresentar data nula, maior ou igual a esta DATA\_REF.
* Envio do parâmetro valido (ID\_OCORRENCIA) ao mapa de controle de finalização (FINALIZA\_CONTROLE).

1. Visão Lógica do Fluxo de Dados



|  |  |  |  |
| --- | --- | --- | --- |
| Fluxo Padrão | Tipo de Carga | Identificador do Fluxo | Atividades |
| Processo de carga das Stages | Incremental | 1 | Fluxo responsável por realizar a recuperação dos dados das tabelas do sistema de origem e enviar para as tabelas Stages. |

* 1. Fluxo Padrão

O fluxo descrito abaixo representa o processo de extração dos sistemas de origem para as Stages. Este processo será incremental ou full, por data de referência.

Inicialmente será executado um mapa de controle inicial (INICIA\_CONTROLE). Este mapa será responsável por definir as informações ID\_OCORRENCIA (identificador da execução) e DATA\_REF (data de referência). Estas informações serão enviadas por parâmetro para o segundo mapa.

O segundo mapa irá realizar a extração das tabelas dos sistemas de origem para as respectivas tabelas de Stage. O mapa receberá, por parâmetro, os valores gerados pelo mapa de controle inicial, realizará a extração dos dados dos sistemas de origem e carregará estes dados nas tabelas de Stage.

Para uma carga full, o parâmetro DATA\_REF deverá ser configurado com um valor inferior ao menor valor de data existente no sistema de origem.

Nas execuções diárias o parâmetro DATA\_REF conterá DATA\_REF deverá conter a data de referência do último processo executado com êxito, recuperando desta forma apenas os novos registros ou atualizações do sistema de origem.

Os registros da origem a serem extraídos, tanto na carga full quanto na incremental, devem apresentar data nula, maior ou igual a esta DATA\_REF

O terceiro mapa (FINALIZA\_CONTROLE) irá atualizar as tabelas de controle de execução com as informações relacionadas a finalização do processo. O único parâmetro necessário para este mapa é o ID\_OCORRENCIA.

Os três passos de execução do processo estarão encapsulados em worklets. Cada worklet terá o nome do processo que ele representa. A relação dos processos que serão replicados encontra-se descrita no documento de solicitação de criação de processos de ETL1 entregue pelo respectivo projeto.

A seguir são apresentados fluxos de exemplo para melhor esclarecimento do acima descrito.



Onde cada worklet é representado da seguinte forma:



|  |  |  |  |
| --- | --- | --- | --- |
| Tarefa | Identificação do Fluxo | Segmentos Simultâneos | Descrição das atividades |
| wf\_carga\_stages\_sia\_to\_bi | 1 | S | Workflow responsável por realizar a extração dos sistemas de origem para as Stages. |

Os mapas receberão os campos P\_DATA\_REF e P\_ID\_OCORRENCIA por parâmetro. Esta data será utilizada no Source Filter das tabelas para realizar a extração incremental dos dados.

A clausula a ser utilizada no Source Filter de cada mapa esta descrita na planilha anexa ao item 3.1.1

* + 1. Estruturas de Dados - m\_alocacao\_extensao



|  |  |  |  |
| --- | --- | --- | --- |
| Tarefa | Identificação do Fluxo | Segmentos Simultâneos | Descrição das atividades |
| DUAL | 1 | N | Tabela padrão do sistema |
| SQ\_DUAL | 2 | N | Source Qualifier da tabela |
| EXP\_GERA\_DADOS | 3 | N | Expression responsável por realizar a chamada da procedure inicial |
| ARQ\_CONTROLE | 4 | N | Arquivo gerado ao final da execução do mapa contendo informações de ID de Ocorrência e Data de referencia |
| PROC\_INICIA | 5 | N | Procedure de banco responsavel por gerar as informações de controle. |



|  |  |  |  |
| --- | --- | --- | --- |
| Tarefa | Identificação do Fluxo | Segmentos Simultâneos | Descrição das atividades |
| SIA\_ALOCACAO\_EXTENSAO | 1 | N | Tabela do sistema SIA onde as informações serão disponibilizadas |
| SQ\_SIA\_ALOCACAO\_EXTENSAO | 2 | N | Source Qualifier, onde o filtro de extração será informado. |
| EXP\_TRATA\_DADOS | 3 | N | Expression responsável por inserir os registros de parâmetros e controle |
| ALOCACAO\_EXTENSAO | 4 | N | Tabela de stage onde os dados serão gravados. |



|  |  |  |  |
| --- | --- | --- | --- |
| Tarefa | Identificação do Fluxo | Segmentos Simultâneos | Descrição das atividades |
| ARQ\_CONTROLE | 1 | N | Arquivo gerado no mapa inicia controle contendo os dados de ID de ocorrência e Data de Referência |
| EXP\_TRATA\_DADOS | 2 | N | Expression responsável por tratar os dados |
| UPD\_CONTROLE | 3 | N | Gera o Update na tabela de controle |
| TABELA\_CONTROLE | 4 | N | Tabela de controle onde os dados serão atualizados |

1. Estratégias e Controles
   1. Controle e Tratamento de Exceções

Descrição das estratégias de controle dos processos e de tratamento de exceções.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Exceção ou Erro | Tipo | Razão | Localização do Log | Ação | Identificação |
| Erro ao ler um registro | Técnico | Layout incorreto, tipo de dados especificados incorretos. | O Log de execução dos processos será disponibilizado na Tabela de controle de carga e Tabela de controle do processo. | Registro do erro no arquivo de log da interface. | Erro gravado no log. |
| Interrupção do processo | Técnico | Interrupção procedente de falha na conexão de alguma das bases de origem ou destino. | O Log de execução dos processos será disponibilizado na Tabela de controle do processo e Tabela de controle de carga. | Interrupção do Processo | Interrupção de Carga |

* 1. Expurgo de Dados

Esta seção deve descrever a estratégia de expurgo de dados em detalhe, determinando o período máximo de retenção e a periodicidade da execução.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Regra | Condição | Tipo | Periodicidade | Período de Retenção |
| Limpa\_Processados | Registros com o status de processamento marcado com o valor ‘S’ | *Delete – 1 block commit;* | Diário | 1 dia. |
| Limpa\_Não\_Processados | Registros com o status de processamento marcado com o valor ‘E’ | *Delete – 1 block commit;* | FIFO | Parametrizável |

* 1. Reinício e Recuperação

Esta seção deve indicar os *workflows* existentes, bem como as seções que devem utilizar as funcionalidades de *Restart* e *Recovery*.

|  |  |  |
| --- | --- | --- |
| Nome do Objeto | Tipo do Objeto | Tipo de Estratégia |
| wf\_carga\_stages\_sia\_to\_bi | *Workflow* | *Resume from last Checkpoint* |

* 1. Envio de Notificações

Para cada requisito de notificação, necessária ao processo, deve ser especificada a descrição do conteúdo da notificação, tipo, momento, e a lista de destinatários para o e envio da notificação e o evento ou condições de disparo.

| Notificação | Tipo | Momento | Destinatário(s) | Gatilho |
| --- | --- | --- | --- | --- |
| O processo referente à data de referência DD/MM/AAAA rodou com falha não concluindo o processamento de: XXXXXX.  Onde:  XXXXXX deve ser a lista de etapas com falha e respectivas informações afetadas. | *E-mail Task* | Ocorrência de erro | Gerente de BI; Gerente de Operações; Gerente de Infraestrutura; Coordenador de Infraestrutura; DBA; Dono da Aplicação; Capacitadores de custo de pessoal; RH. | Interrupção de Carga |
| Os processos abaixo referentes à dada de referência DD/MM/AAAA rodaram com êxito, porém apresento registros com erros ou não processados de acordo com o log anexo:  XXXXXX apresentou YY registros com erros ou não processados de um total de TT.  Onde:   * XXXXXX deve ser a lista de etapas com falha e respectivas informações afetadas, tendo cada etapa:   + o número (YY) de registros com erros ou não processados;   e o total de registros TT. | *E-mail Task* | Término da execução | Gerente de BI; Gerente de Operações; DBA; Capacitadores de custo de pessoal; RH. | Erro ao ler um registro |

1. Definição de Abreviaturas, Siglas e Acrônimos

SIA – Sistema de Informação Acadêmica.

RNx – Requisito não funcional de número x

RFx – Requisito funcional de número x

[Dê uma visão geral das funcionalidades para a área de Negócio]

[Descrever de forma objetiva os requisitos ou a essência das funcionalidades para este projeto]

[Esta parágrafo serve para informar detalhes financeiros do projeto. Em particular, definições de budget do projeto, entretanto se existir qualquer ferramenta ou funcionalidade em particular que necessite de algum budget específico deverá ser informado nesta seção..]