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

Extração de Área de Estágio e Cargas na ODS

Versão 1.1

10/12/2012

Histórico de Revisões

|  |  |  |  |
| --- | --- | --- | --- |
| Data | Versão | Preparado por | Comentários |
| 10/12/2012 | 1.0 | Vanessa Felix/ Hugo Cardoso | Versão Inicial do documento |
| 17/12/2012 | 1.1 | Hugo Ayres Cardoso | Alteração conforme reunião de validação dia 14/12/2012. |

Índice

[Histórico de Revisões 1](#_Toc343589492)

[1 Objetivo 2](#_Toc343589493)

[2 Escopo 3](#_Toc343589494)

[2.1 Descrição do Escopo 3](#_Toc343589495)

[2.1.1 Requisitos Funcionais 3](#_Toc343589496)

[2.1.2 Requisitos Não Funcionais 3](#_Toc343589497)

[2.2 Fora do Escopo 4](#_Toc343589498)

[2.3 Premissas 4](#_Toc343589499)

[3 Visão Lógica do Fluxo de Dados 4](#_Toc343589500)

[3.1 Fluxo Padrão 5](#_Toc343589501)

[3.1.1 Estruturas de Dados - m\_alocacao\_extensao 6](#_Toc343589502)

[4 Estratégias e Controles 10](#_Toc343589504)

[4.1 Controle e Tratamento de Exceções 10](#_Toc343589505)

[4.2 Expurgo de Dados 10](#_Toc343589506)

[4.3 Reinício e Recuperação 11](#_Toc343589507)

[4.4 Envio de Notificações 11](#_Toc343589508)

[4.5 Reprocessamento 12](#_Toc343589509)

[5 Definição de Abreviaturas, Siglas e Acrônimos 13](#_Toc343589510)

1. Objetivo

O documento atual tem como objetivo descrever o processo padrão de coleta das informações das tabelas de Stage para carga no BI da Estácio. Neste processo as informações serão extraídas de forma full das tabelas de stage, realiza uma comparação com os dados já existentes na ODS e, caso o registro não exista este é inserido, caso ele já exista é criada uma nova versão deste registro.

1. Escopo
   1. Descrição do Escopo

Este documento tem como objetivo estabelecer as definições padrões para desenvolvimento das interfaces que realizarão extrações das tabelas de stage e enviarão estes dados para as suas respectivas tabelas de ODS. Este processo será realizado para todas as tabelas envolvidas nos processos de ETL 2 ( STAGE -> ODS) da arquitetura de BI.

O processo de criação das tabelas de ODS utilizará, quando disponível, o script das tabelas do sistema legado modificado para conter os mesmos atributos e datatypes, inclusive as informações de chave primária e chave estrangeira. Alem dos respectivos campos de origem, serão acrescidos 6 novos campos a estas tabelas. São eles:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nome do campo | Tipo | Descrição do campo | Permite Nulos | Regra de Preenchimento |
| ID\_OCORRENCIA | LONG INT | Identificador da execução do processo | Não | P\_ID\_OCORRENCIA |
| DT\_INC\_REGISTRO | DATE | Data de inclusão do registro utilizando a data de referência da cadeia. | Não | Rastreamento de reprocessamento |
| DT\_ALT\_REGISTRO | DATE | Data de alteração do registro utilizando a data de referência da cadeia. | Não | Rastreamento de reprocessamento |
| ORIGEM\_INFORMACAO | VARCHAR2(20) | Indica o sistema fonte do registro | Não | Sigla do sistema de origem. Ex: SIA, Manual, CSV, ADP, SAP, .. |
| DT\_INI\_VIGENCIA | DATE | Data de inicio de vigência utilizando a data de inclusão do registro no legado, caso esteja nulo será inserido ‘1/1/1900’. | Não |  |
| DT\_FIM\_VIGENCIA | DATE | Data de fim de Vigência. | Não |  |
| IND\_ATIVO\_LOG | INT | Indicador do status do registro (1 – Ativo; 0 – Inativo; 2 – Deleção Lógica). | Não |  |

A relação contendo o nome de todas as tabelas 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 ETL2.

* + 1. Requisitos Funcionais

Este documento faz referencia a regra funcional RF35 do documento Requisitos Funcionais.xls

* + 1. Requisitos Não Funcionais

N/A

* 1. Fora do Escopo

N/A

* 1. Premissas
* A cadeia referente a carga das stages devem estar concluídas com sucesso.
* Execução com êxito do respectivo mapa de controle inicial (INICIA\_CONTROLE);
* Recebimento de parâmetros validos com as informações ID\_OCORRENCIA (identificador da execução) ;
* 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 da ODS | Insert/  Update | 1 | Fluxo responsável por realizar a recuperação dos dados das tabelas Stages e enviá-los as tabelas da ODS |

* 1. Fluxo Padrão

O fluxo descrito abaixo representa o processo de extração da stage para o modelo ODS. Este processo será de Insert/Update e conterá três mapas.

Inicialmente, será executado um mapa de controle inicial (INICIA\_CONTROLE). Este mapa será responsável por definir a informação de ID\_OCORRENCIA (identificador da execução). Esta informação será enviada, por parâmetro, para o mapa principal.

O segundo mapa irá realizar a extração das tabelas stage para as respectivas tabelas da ODS. O mapa receberá, por parâmetro, o valor gerado pelo mapa primeiro mapa e realizará a extração dos dados da stage, carregando estes dados nas tabelas de ODS.

O processo de carga da tabela ODS consiste em extrair os dados das tabelas do ambiente de stage realizando uma comparação com os dados já existentes na ODS por left join no Source Qualifier (referente a ODS.ID = STG.ID (+)). Caso o registro não exista, este será inserido com o campo **DT\_INICIO\_VIGENCIA** igual o campo data de inclusão do registro no legado, se o campo for nulo, deverá ser preenchido com o valor do campo DT\_ATUALIZA\_LOG pertencente ao legado, e o campo **DT\_FIM\_VIGENCIA** será preenchido, por padrão, com o valor ***‘31/12/2999’***; O campo **IND\_ATIVO\_LOG** será preenchido com o valor inicial ***1*** (ativo). As datas **DT\_INC\_REGISTRO** e **DT\_ALT\_REGISTRO** deverão ser preenchidas com a **DATA\_REF** da cadeia em andamento. Quando o registro foi alterado no legado, este registro será atualizado. Assim, o preenchimento do campo **DT\_FIM\_VIGENCIA** passará a ser ***‘DATA\_REF -1’*** e o campo **IND\_ATIVO\_LOG** será atualizado para ***0*** (inativo) e atribuindo a data **DT\_ALT\_REGISTRO** o novo valor de **DATA\_REF** da cadeia em execução. Uma nova linha será inserida na tabela com o padrão de vigência ativa conforme passo anterior, mas com uma diferença, atribuir **DATA\_REF a DT\_INICIO\_VIGENCIA**, indicando que este registro é o mais atual da tabela.

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 ETL2 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\_load\_bi\_ods\_custo\_pessoal | 1 | S | Workflow responsável por realizar a extração das tabelas de stage para as tabelas da ODS. |

* + 1. Estruturas de Dados - m\_load\_bi\_ods\_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 responsável por gerar as informações de controle. |



|  |  |  |  |
| --- | --- | --- | --- |
| Tarefa | Identificação do Fluxo | Segmentos Simultâneos | Descrição das atividades |
| STAGE\_ALOCACAO\_EXTENSAO | 1 | S | Tabela de stage onde os dados serão disponibilizados |
| SQ\_STAGE\_ALOCACAO\_EXTENSAO | 2 | N | Source Qualifier responsável pela verificação de vigência dos registros, onde os dados da STG são mandatórios. Realizando um left join de referência ODS.ID = STG.ID (+). |
| ODS\_ALOCACAO\_EXTENSAO | 3 | S | Tabela ODS onde os dados serão disponibilizados |
| EXP\_INS\_REG\_CTRL | 4 | N | Expression responsável por inserir os registros de parâmetros e controle. |
| RTR\_CONTROLA\_REGISTRO | 5 | N | Router responsável direcionar os registros novos e os alterados. Se o ID d ODS for NULO, representa os registro novos, caso contrario são antigos. |
| INS\_NOVO\_ODS\_ALOC\_EXT | 6 | N | Área de insert dos registros novos. |
| UPD\_ATUALIZA\_ANTIGO | 7 | N | Update strategy responsável pela atualização dos dados existentes no ODS. |
| UPD\_ANTIGO\_ODS\_ALOC\_EXT | 8 | N | Área de update dos registros antigos que serão inativados. |
| INS\_ANTIGO\_ODS\_ALOC\_EXT | 9 | N | Área de insert dos registros antigos atualizados, que serão ativados. |



|  |  |  |  |
| --- | --- | --- | --- |
| 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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Exceção ou Erro | Tipo | Razão | Localização do Log | Ação | Identificação |
| Interrupção do processo | Técnico | Layout das tabelas de origem ou destino incorreto, interrupção procedente de falha na conexão de alguma das bases de origem ou destino, etc. | Tais registros estarão disponibilizados no arquivo de extensão \*.log na pasta padrão de logfiles do Informatica com o seguinte padrão: <nome da session>.log. | Interrupção do Processo e registro do erro no arquivo de log da interface. | Interrupção de Carga e erro gravado no arquivo de log. |
| Erro ao inserir um registro | Técnico | Registros rejeitados por duplicidade ou tipo de dados especificados incorretos. | Tais registros estarão disponibilizados no arquivo de extensão \*.bad na pasta padrão de badfiles do Informatica com o seguinte padrão: <nome da session>.bad. | O processo ficará Interrompido até que se investigue o motivo de rejeição. | Interrupção da Carga |

* 1. Expurgo de Dados

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Regra | Condição | Tipo | Periodicidade | Período de Retenção |
| Limpa\_ODS | Registros retidos nas tabelas do ODS com mais de 5 anos. | * Delete where DATE\_DIFF($$DATA\_REF - DATA\_REF, ‘YYYY’) > 5   Commit; | Sugestão: execução Trimestral | 5 anos. |

* 1. Reinício e Recuperação

|  |  |  |
| --- | --- | --- |
| Nome do Objeto | Tipo do Objeto | Tipo de Estratégia |
| wf\_load\_bi\_ods\_custo\_pessoal | Workflow | Restart Workflow from Task |

* 1. Envio de Notificações

Inserir aonde será configurado essas regras.

| Notificação | Tipo | Momento | Destinatário(s) | Gatilho |
| --- | --- | --- | --- | --- |
| O processo referente à data de execução <***Session Start Time>*** rodou com falha, não concluindo o processamento de: <***Session Name>***. Apresentando ***<Total Records regected>*** registros com erros ou não processados de um total de ***<Total Records regected> + <Total Records loaded>***. | *Session 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 por erro |
| O processo referente à data de execução <***Session Start Time>*** rodou com êxito, concluindo o processamento de <***Session Name>***. Apresentando um total de registros ***<Total Records loaded>*** carregados. | *Session E-mail Task* | Ocorrência de sucesso | Gerente de BI; Gerente de Operações; Gerente de Infraestrutura; Coordenador de Infraestrutura; DBA; Dono da Aplicação; Capacitadores de custo de pessoal; RH. | Carga executada com sucesso |

* 1. Reprocessamento

Quando houver necessidade de reprocessamento, deverá ser feito uma deleção lógica dos registros do ODS a partir da data de referência da cadeia ***(DATA\_REF)*** de reprocessado. Pois se os registros que compõem a stage, não tiverem mais no ODS, a partir do período extraído do legado, a ***DT\_FIM\_VIGENCIA*** deverá ser atualizada com ***DATA\_REF – 1***, o campo ***IND\_ATIVO\_LOG*** alterado para o status **‘2’** ea data **DT\_ALT\_REGISTRO** preenchida com a **DATA\_REF** da cadeia em reprocessamento.

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..]