

OBJETIVO

Describir los requerimientos y actividades necesarias para realizar el aseguramiento de la calidad de los datos en la Bodega de datos, con el fin de generar información confiable para el análisis de datos y toma de decisiones de temas relacionados con Seguridad, Convivencia y Justicia.

GLOSARIO

Bodega de datos: Herramienta empresarial que consolida los datos de diferentes fuentes de información, ya sean de bases de datos, archivos planos u otros sistemas, extrayéndolos, transformarlos y almacenándolos en un solo repositorio.

ETL (Extract, Transform, Load): Proceso de compilación de datos a partir de un número de fuentes, posterior compilación y almacenamiento en un único repositorio.

OAIEE: Oficina de Análisis de Información y Estudios Estratégicos.

SACD: Sistema de Aseguramiento de Calidad de Datos.

RESPONSABLES

Oficina de Análisis de Información y Estudios Estratégicos.

DESARROLLO

ASEGURAMIENTO DE CALIDAD DE DATOS

Inicia con el diligenciamiento del formato de especificaciones por parte del área funcional, continua con el análisis de las reglas de validación y la traducción informática y finaliza con la ejecución y generación de resultados por parte de la Oficina de Análisis de Información y Estudios Estratégicos. Las actividades descritas en el presente documento aplican a fuentes de datos que permitan hacer modificaciones o ajustes a los elementos propios de la base de datos como campos, tablas o registros.

1. ASEGURAMIENTO DE CALIDAD DE DATOS – SACD

1.1. Diseño Técnico del Proceso

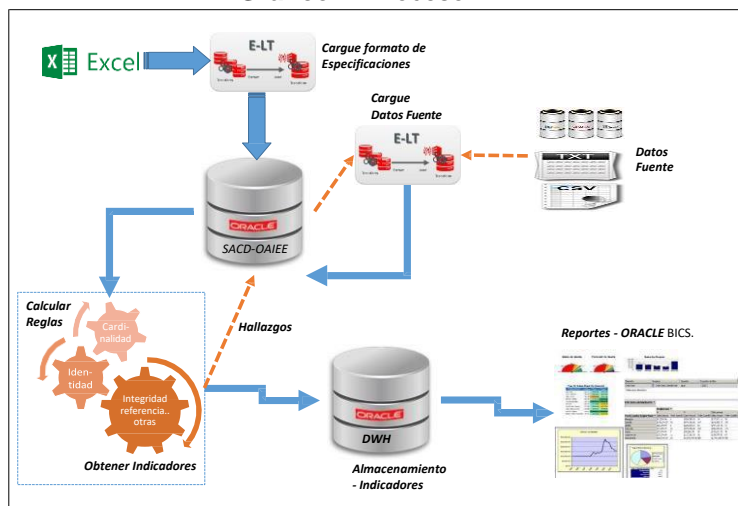
El sistema de aseguramiento de calidad de datos, está conformado por una serie de pasos ordenados que se deben realizar para la obtención de un indicador que permita identificar la calidad de los datos que se almacenan en una base de datos específica.

El diseño requiere una entrada de datos, que corresponde a un archivo en Excel denominado formato de especificaciones; un proceso de análisis del formato de especificaciones y traducción de reglas de integridad de lenguaje natural a lenguaje SQL; un proceso ETL que orquesta los pasos para

DESARROLLO

realizar el cargue del formato de especificaciones a una base de datos relacional; obtención de datos fuente; procesamiento y generación de resultados.

Gráfico 1. Proceso ETL



1.2. Diligenciamiento Archivo Excel Formato de especificaciones

El formato de entrada de datos, corresponde a un archivo en formato Excel, que contiene cada uno de los campos requeridos por el sistema para identificar de manera única cada una de las operaciones sujetas al proceso de análisis de calidad, sus archivos o tablas fuentes y cada uno de los campos que la componen, a los que se aplican reglas de validación y consistencia.

El archivo Excel, que fue diseñado, consta del Instructivo Diligenciamiento; datos de la entidad; Datos de la operación; Atributos por archivo o tabla fuente; y una última hoja para realizar la identificación de la regla de validación asociada a cada campo y su correspondiente traducción informática que estará a cargo de una analista de la OAIEE.

Ver Formato: “SACD_FTE_Formato_EspecificacionesV2.xlsx”.

1.3. Revisión formato de especificaciones Fuentes de Información

Este proceso es realizado por el administrador del SACD, el cual debe revisar el correcto diligenciamiento del formato de especificaciones, identificando que las tabla incluidas en la *Hoja “3. Fuentes”* del formato de especificaciones si se encuentren con el nombre y en el schema de base de datos del sistema fuente. En la Hoja “3.1. Atributos” se deben revisar que las tablas definidas coincidan con las de la anterior (Hoja “3. Fuentes”) que los campos de las tablas existan; identificar si se aplicaron llaves primarias, foráneas; que los dominio o listas de valores estén definidos acorde al formato requerido; que las reglas en lenguaje natural sean lo suficientemente explícitas y que le

DESARROLLO

permitan identificar claramente cómo se debe realizar la traducción informática de cada regla a lenguaje SQL.

Cumplidas las condiciones anteriores se puede proceder a realizar la traducción o en caso contrario se debe realizar una revisión en cada una de las reglas que se consideran incompletas con el fin de aclararlas con el funcional o parte técnica del sistema de información.

1.3.1. Parametrización “cargar tabla”

Cada vez se defina una nueva tabla fuente en el formato de especificaciones; *Hoja “3. Fuentes”*; es necesario incluirla también en la tabla de control “STG_TBL_CONTROL_CARGA”. Para esto se debe obtener el nombre de la tabla final (tabla DAT_) que se conforma con la siguiente regla, iniciando con el prefijo DAT_; seguido por las iniciales de la entidad, seguido de la sigla de la operación, y el nombre de la tabla original sin espacios ni caracteres especiales; hasta completar un máximo de 30 caracteres.

DAT_ENTTOPPFUENTE

ENTT: Corresponde a la sigla de la entidad fuente. OPP: Sigla de la operación a aplicativo a evaluar. FUENTE: Nombre de la tabla o archivo fuente.

Por ejemplo, para el caso de SICAS, se definieron los siguientes nombres de tablas temporales:

Gráfico 2. Ejemplo Tabla Temporal

STR_NOM_TABLA_FUENTE_NK	STR_NOM_TABLA_TEMPORAL
CASA JUSTICIA	DAT_SCJSCASCASAJUSTICIA
LOCALIDAD	DAT_SCJSCASLOCALIDAD
NACIONALIDAD	DAT_SCJSCASNACIONALIDAD
PERMISO SICAS	DAT_SCJSCASPERMISOSICAS
ESTRUCTURA_FAMILIAR	DAT_SCJSCASESTRUCTURAFAMILIAR
COMO_SE_ENTERO	DAT_SCJSCASCOMOSEENTERO
BARRIO	DAT_SCJSCASBARRIO
HABITANTE	DAT_SCJSCASHABITANTE
PROBLEMA_CRI	DAT_SCJSCASPROBLEMACRI
TIPO_DOCUMENTO	DAT_SCJSCASTIPODOCUMENTO
CARACTERIZACION_CRI	DAT_SCJSCASCARACTERIZACIONCRI
PROBLEMA_CARACTER...	DAT_SCJSCASPROBLEMACARACTERIZ
VISITA_CASA JUSTICIA	DAT_SCJSCASVISITACASAJUSTICIA
CIUDAD	DAT_SCJSCASCIUDAD
DEPARTAMENTO	DAT_SCJSCASDEPARTAMENTO
ENTERO	DAT_SCJSCASEENTERO
FUNCIONARIO_CASA	DAT_SCJSCASFUNCIONARIOCASA
DIRIGIDO_DESDE_CRI	DAT_SCJSCASDIRIGIDODESDECRI

1.3.2. Traducción informática de reglas de validación

Este paso consiste en tomar cada una de las reglas definidas en lenguaje natural y realizarle su correspondiente traducción a lenguaje SQL; referenciando siempre en la parte del FROM el nombre de la tabla de trabajo “DAT_”; y no el de la tabla original.

DESARROLLO

Una vez traducidas cada una de las reglas de forma separada, se deben probar en el sistema fuente, garantizando que el SQL ejecute correctamente, y que no van a generar errores durante el proceso de ejecución.

Cada una de las reglas se deben registrar en la columna “TRADUCCION INFORMATICA” relacionada en la Hoja “4. IdentificaReglas”. del formato de especificaciones.

1.4. Proceso de Cargue Sacd

1.4.1. Preparación de archivos

Cuando el contenido del formato de especificaciones sea aceptado en su totalidad por el administrador del SACD, y se hallan incluido todas las reglas en lenguaje natural, que se tratan en el punto “INTRODUCCIONRevisión formato de especificaciones”, se debe disponer el archivo Excel en la siguiente ruta, desde donde se alimentará el sistema para cargar cada una de sus hojas al modelo relacional.

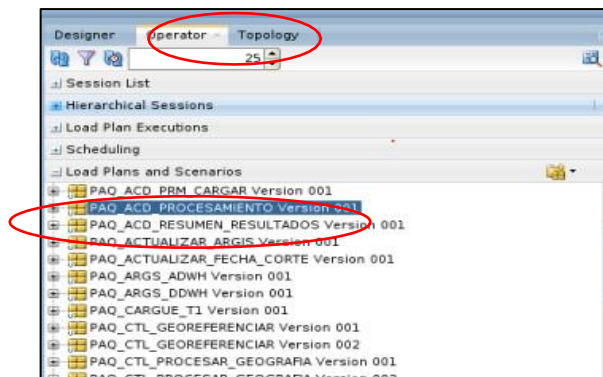
Tabla 1. Archivos fuentes utilizados para el cargue de SIEDCO

CARPETA –FTP	Archivo Formato de Especificaciones
/u01/Dataserver/SACD/1.PLANTILLAS/	Formato_Especificaciones_SCJ_SEDCO.xlsx

1.4.2. Ejecución del Programa Principal

El Scenario principal que orquesta todo el proceso de cargue de SACD, es “PAQ_ACD_PROCESAMIENTO”. Está ubicado en la pestaña “Operator”, y en la sección “Load Plans and Scenarios”:

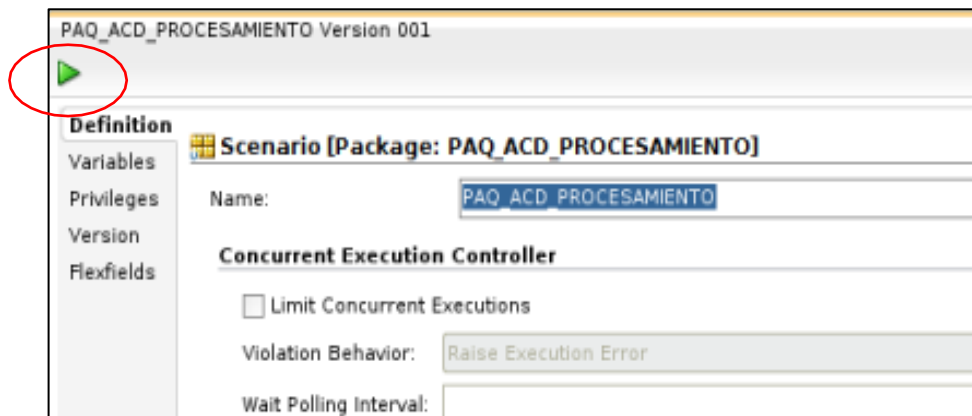
Gráfico 3. Ubicación del paquete principal



DESARROLLO

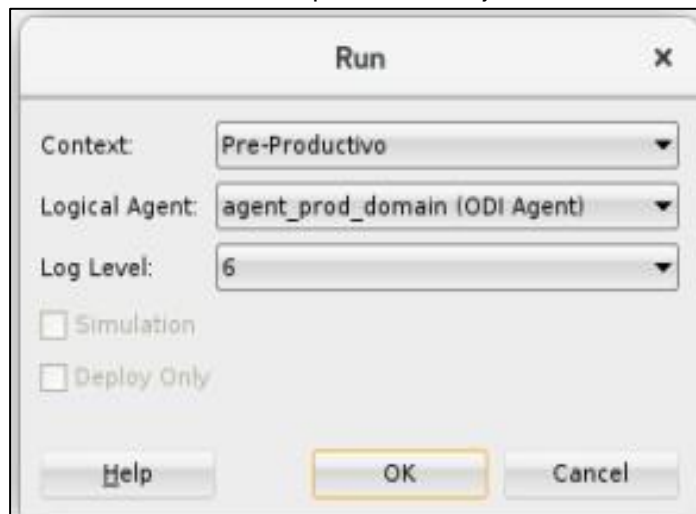
Para ejecutarlo, seleccione el escenario y con clic derecho elija la opción "Run" o ábralo con doble clic y ejecútelo desde el icono verde.

Gráfico 4. Paso de ejecución



Se debe ejecutar en el contexto Pre-Productivo, con el agente y nivel de log 6, como se indica en la imagen, luego OK.

Gráfico 5. Propiedades de Ejecución

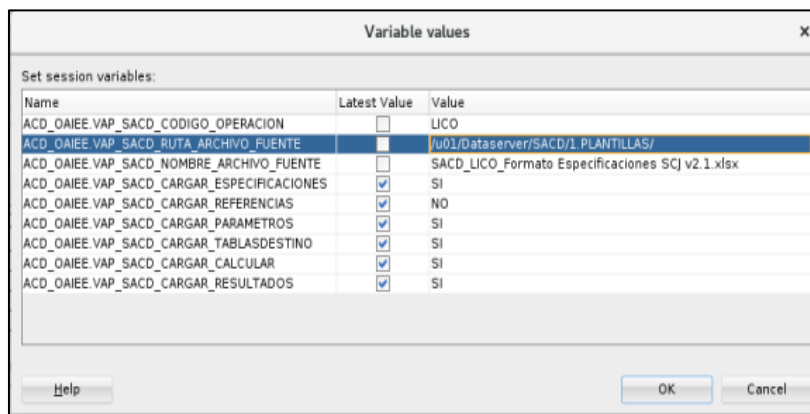


Parámetros de entrada

Dependiendo del valor indicado en los parámetros; se pueden realizar u omitir algunos pasos en el proceso de carga:

DESARROLLO

Gráfico 6. Parámetros de entrada Paquete Principal



Name	Latest Value	Value
ACD_OAIEE.VAP_SACD_CODIGO_OPERACION	<input type="checkbox"/>	LICO
ACD_OAIEE.VAP_SACD_RUTA_ARCHIVO_FUENTE	<input type="checkbox"/>	/u01/Dataserver/SACD/1.PLANTILLAS/
ACD_OAIEE.VAP_SACD_NOMBRE_ARCHIVO_FUENTE	<input type="checkbox"/>	SACD_LICO_Formato Especificaciones SCJ v2.1.xlsx
ACD_OAIEE.VAP_SACD_CARGAR_ESPECIFICACIONES	<input checked="" type="checkbox"/>	SI
ACD_OAIEE.VAP_SACD_CARGAR_REFERENCIAS	<input type="checkbox"/>	NO
ACD_OAIEE.VAP_SACD_CARGAR_PARAMETROS	<input checked="" type="checkbox"/>	SI
ACD_OAIEE.VAP_SACD_CARGAR_TABLASDESTINO	<input checked="" type="checkbox"/>	SI
ACD_OAIEE.VAP_SACD_CARGAR_CALCULAR	<input checked="" type="checkbox"/>	SI
ACD_OAIEE.VAP_SACD_CARGAR_RESULTADOS	<input checked="" type="checkbox"/>	SI

Lista de parámetros de entrada que requiere el paquete, para la ejecución, cargue y procesamiento del indicador. Los valores indicados en Descripción son ejemplos, dependiendo de la fuente a realizar análisis, deberán cambiar a los valores correspondientes.

Tabla 2. Listado de parámetros paquete principal

Nombre del parámetro	Tipo de dato	Descripción
ACD_OAIEE.VAP_SACD_CODIGO_OPERACION	Texto	LICO
ACD_OAIEE.VAP_SACD_RUTA_ARCHIVO_FUENTE	Texto	Es la ruta del servidor ODI donde se encuentra ubicado el archivo. Ejemplo: /u01/Dataserver/SACD/1.PLANTILLAS/
VAP_SACD_NOMBRE_ARCHIVO_FUENTE	Texto	El nombre del formato de entrada de datos. Ejemplo: SACD_LICO_Formato Especificaciones SCJ v2.1.xlsx
ACD_OAIEE.VAP_SACD_CARGAR_ESPECIFICACIONES	Texto	SI: se cargará el formato de especificaciones definido con el nombre de archivo y ruta fuente definidos en las variables anteriores. NO: No se realiza el cargue del formato y se mantendrá la última estructura y reglas cargadas.
ACD_OAIEE.VAP_SACD_CARGAR_REFERENCIAS	Texto	SI: Indica que se van a cargar las tablas de referencia. Esto solo se hace una sola vez; y no es necesario volver a realizarlo a menos que se quieran refrescar los valores en el sistema relacional. NO: No se realiza el cargue de tablas de referencia. (Valor por defecto NO).

DESARROLLO			
ACD_OAIEE.VAP_SACD_CARGAR_PARAMETROS	Texto	<p>SI: Indica que se van a cargar las tablas de parámetros, al sistema relacional; es decir los datos del formato de especificaciones se cargan al sistema relacional, generando llaves únicas, para cada uno de las variables y tablas definidas en el formato.</p> <p>NO: No se realiza el cargue de tablas de parametros. (Valor por defecto SI).</p>	
ACD_OAIEE.VAP_SACD_CARGAR_TABLASDESTINO	Texto	<p>SI: se cargaran cada una de las tablas del sistema fuente a su correspondiente tabla DAT_ENTTOPPFUENTE</p> <p>NO: No se realiza el cargue de datos fuente y se mantendrán los últimos datos cargados.</p>	
ACD_OAIEE.VAP_SACD_CARGAR_CALCULAR	Texto	<p>SI: Indica que se va a realizar el cálculo del indicador con las nuevas reglas que se carguen del formato de especificaciones.</p> <p>NO: No se realiza proceso de cálculo.</p>	
ACD_OAIEE.VAP_SACD_CARGAR_RESULTADOS		<p>SI: Indica que se deben generar los archivos con las salidas de información, del indicador y los registros inválidos.</p> <p>Los archivos son dispuestos en la siguiente ruta del FTP:</p> <p>/u01/Dataserver/SACD/3.RESULTADOS 1.SACD_RESUMEN_INDICADOR_LICO_20200306.xlsx 2.SACD_REGISTROS_INVALIDOS_LICO_20200306.xlsx</p>	

1.5. Manual Técnico - Sistema de Aseguramiento de Calidad

1.5.1. Cargue de archivo fuente

El sistema de aseguramiento de calidad de datos, inicialmente realizar el cargue del formato de entrada de datos "Formato_Especificaciones_SCJ.xlsx", que corresponde a un archivo en formato Excel, que contiene cada uno de los campos requeridos por el sistema para identificar de manera única cada una de las operaciones sujetas al proceso de análisis de calidad, sus archivos o tablas fuentes y cada uno de los campos que la componen, a los que se aplican reglas de validación y consistencia.

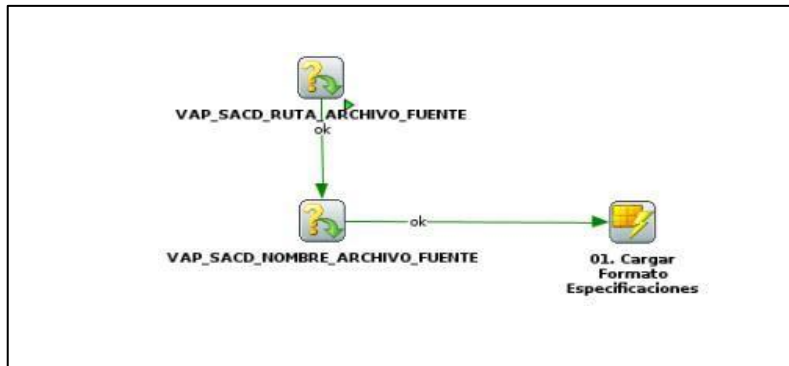
Cada hoja de este archivo se lleva su correspondiente tabla de trabajo del modelo SACD; identificadas por el prefijo "ACD_FRM". Estas tablas se truncan y posteriormente se cargan con los datos proporcionados en el archivo; que luego se transforman y cargan en el conjunto de tablas de parametrización identificadas por el prefijo "ACD_PRM". Ver Formato "Especificaciones Fuente de Información".

Este cargue se realiza desde ODI, a partir del PAQ_ACD_CARGAR_FORMATO_ESPECIFICACIONES

DESARROLLO

i) PAQ_ACD_CARGAR_FORMATO_ESPECIFICACIONES

Gráfico 7. Cargue de parametrización



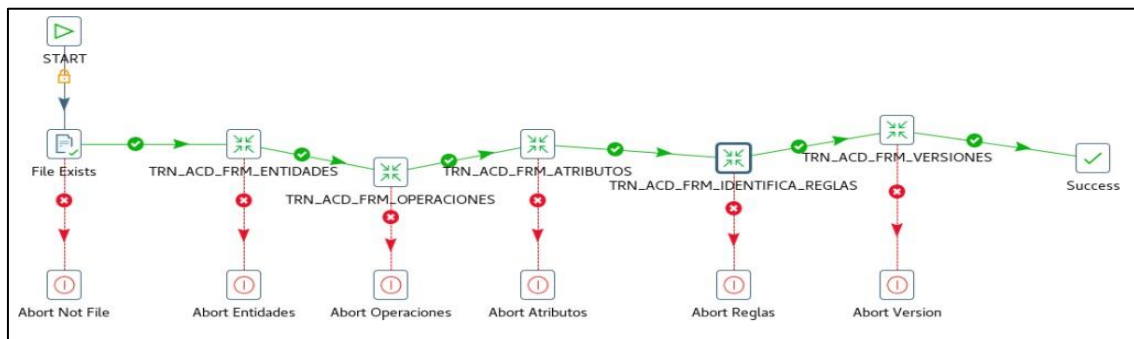
Invoca el JOB de PDI “JOB_ACD_CARGUE_PARAMETRIZACION”, que realiza el cargue del formato de entrada de datos.

Solicita los siguientes parámetros:

- VAP_SACD_RUTA_ARCHIVO_FUENTE: Es la ruta del servidor ODI donde se encuentra ubicado el archivo. Ejemplo: /u01/Dataserver/SACD/1.PLANTILLAS/
- VAP_SACD_NOMBRE_ARCHIVO_FUENTE: El nombre del formato de entrada de datos. Ejemplo: Formato_Especificaciones_SCJ-SIEDCO.xlsx

o **JOB_ACD_CARGUE_PARAMETRIZACION**

Gráfico 8. JOB_ACD_CARGUE_PARAMETRIZACIÓN



DESARROLLO

Este Job, orquesta el proceso de cargue de cada una de las hojas del archivo Excel formato de entrada de datos. Formato_Especificaciones_SCJ-XXXXX.xlsx, para ser cargadas a su tabla de base de datos correspondiente en el SACD.

Recibe los parámetros

- P_VAR_RUTA: VAP_SACD_RUTA_ARCHIVO_FUENTE:
- P_VAR_ARCHIVO: VAP_SACD_NOMBRE_ARCHIVO_FUENTE:

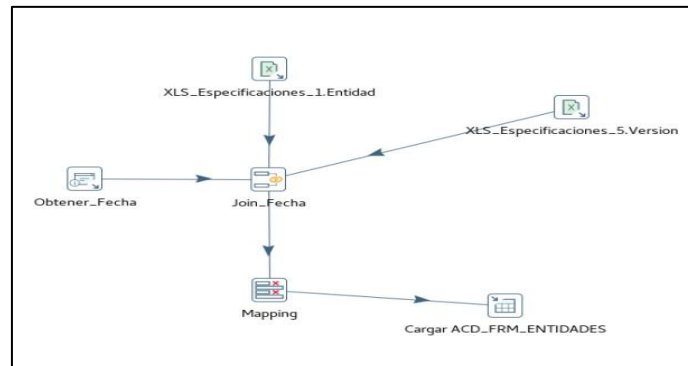
Los parámetros recibidos en este JOB, también son enviados a cada una de las transformaciones utilizadas para el cargue de cada una de las tablas de formulario ACD_FRM_XXX.

○ TRN_ACD_FRM_ENTIDADES

Proceso de cargue de la hoja de datos “1.Entidad” donde registra la información de la entidad o área responsable de la información a ser evaluada:

El proceso obtiene los datos del archivo Excel, de la hoja “1.Entidad” ACD_FRM_ENTIDADES.

Gráfico 9. TRN_ACD_FRM_ENTIDADES

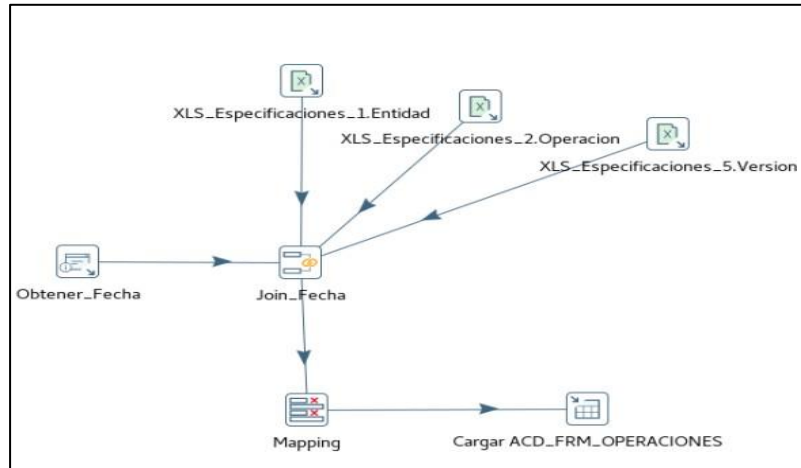


○ TRN_ACD_FRM_OPERACIONES

El proceso obtiene los datos del archivo Excel, de la hoja “2.Operation” donde se encuentran los datos de la operación a ser evaluada, y son cargados en la tabla ACD_FRM_OPERACIONES.

DESARROLLO

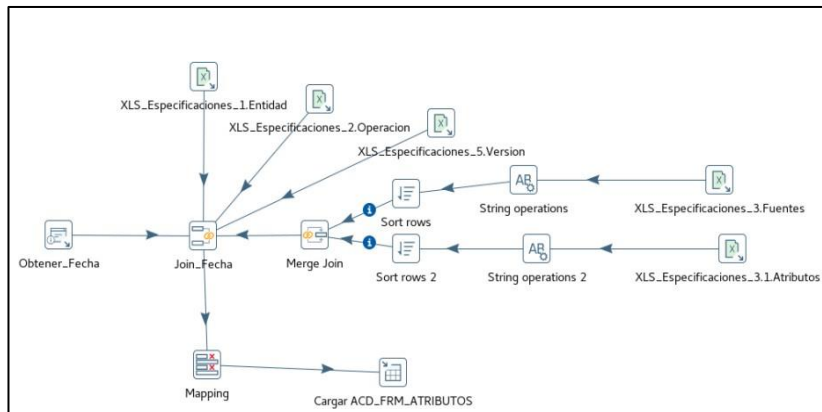
Gráfico 10. TRN_ACD_FRM_OPERACIONES



○ **TRN_ACD_FRM_ATRIBUTOS**

Proceso que extrae los datos del archivo Excel, de la hojas “3.Fuentes” y “3.1.Atributos” donde se encuentran cada uno de los campos de los archivos o tablas fuentes reportados y son cargados en la tabla ACD_FRM_ATRIBUTOS.

Gráfico 11. TRN_ACD_FRM_ATRIBUTOS

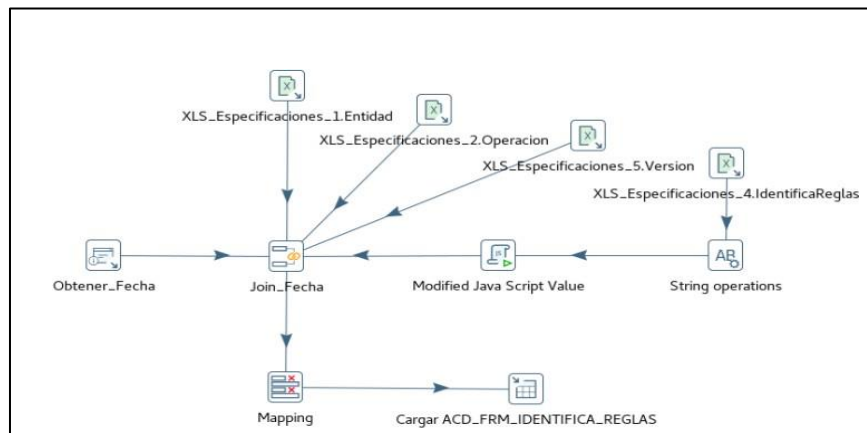


DESARROLLO

○ **TRN_ACD_FRM_IDENTIFICA_REGLAS**

El proceso extrae los datos del archivo Excel, de la hoja “4. IdentificaReglas” donde se están cada una de las reglas que se deben aplicar sobre los archivos o tablas fuentes de la operación, y que son cargados en la tabla ACD_FRM_IDENTIFICA_REGLAS.

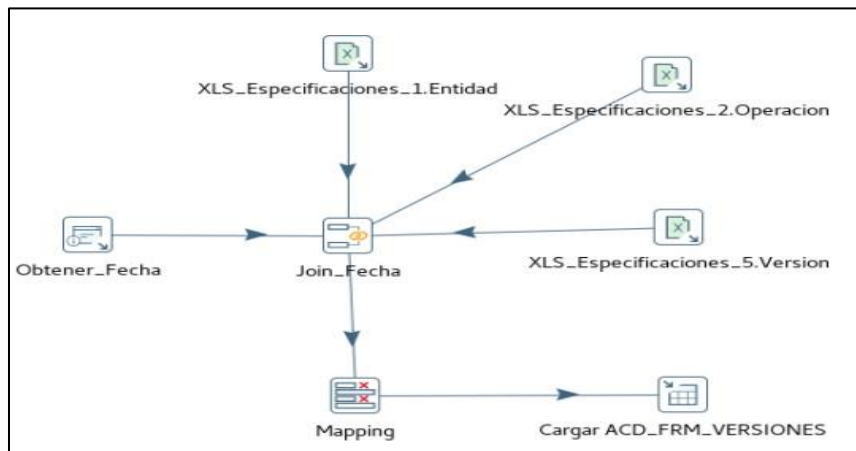
Gráfico 12. TRN_ACD_FRM_IDENTIFICA_REGLAS



○ **TRN_ACD_FRM_VERSIONES**

Realiza la extracción de la versión del archivo de especificaciones, para cargarla sobre la tabla ACD_FRM_VERSIONES.

Gráfico 13. TRN_ACD_FRM_VERSIONES



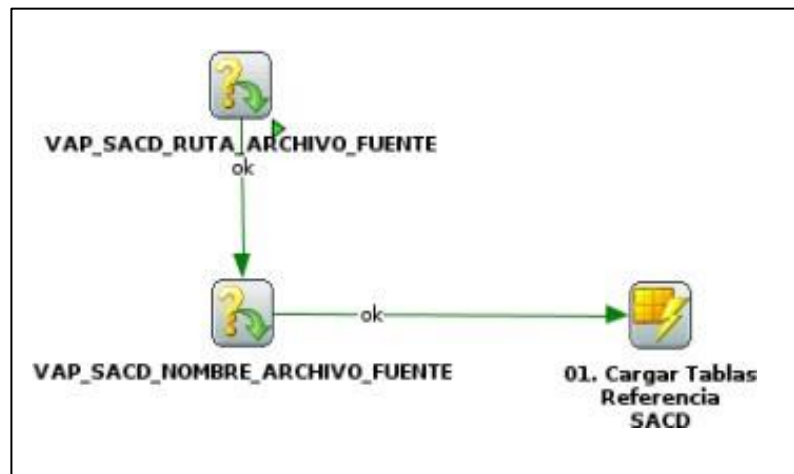
DESARROLLO

1.5.2. Cargue de tablas de referencia

Este proceso se ejecuta a partir del paquete de ODI PAQ_ACD_CARGAR_REFERENCIAS, que a su vez invoca el job de PDI “JOB_ACD_CARGUE_REFERENCIAS”, para realizar el cargue de las tablas de referencia, como tipos de datos; categorías de las reglas, tipos de reglas, reglas entre otras. Utilizadas durante el proceso de calificación y certificación de los datos. Se debe invocar una sola vez.

i) PAQ_ACD_CARGAR_REFERENCIAS

Gráfico 14. PAQ_ACD_CARGAR_REFERENCIAS



Invoca el JOB de PDI “JOB_ACD_CARGUE_REFERENCIAS”.

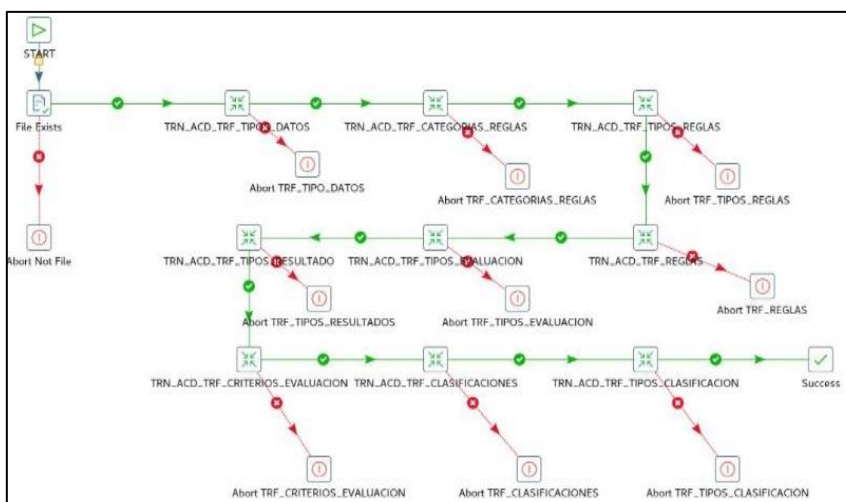
Solicita los siguientes parámetros:

- VAP_SACD_RUTA_ARCHIVO_FUENTE: Es la ruta del servidor ODI donde se encuentra ubicado el archivo. Ejemplo: /u01/Dataserver/SACD/1.PLANTILLAS/
- VAP_SACD_NOMBRE_ARCHIVO_FUENTE: El nombre del formato de entrada de datos. Ejemplo: Formato_Especificaciones_SCJ-SIEDCO.xlsx

DESARROLLO

○ **JOB_ACD_CARGUE_REFERENCIAS**

Gráfico 15. JOB_ACD_CARGUE_REFERENCIAS



Su función es realizar el cargue de cada una de las tablas de referencia “ACD_TRF_xxx”, a partir de las hojas que contienen los datos de referencia del SACSD. Este proceso solo se debe hacer una sola vez, y no es necesario volver a cargarlo, salvo que se definan nuevos dominios a nivel de la parametrización utilizada para el diligenciamiento del formato de especificaciones.

Solicita los siguientes parámetros solicitados por el JOB:

- P_VAR_RUTA: VAP_SACD_RUTA_ARCHIVO_FUENTE
- P_VAR_ARCHIVO: VAP_SACD_NOMBRE_ARCHIVO_FUENTE

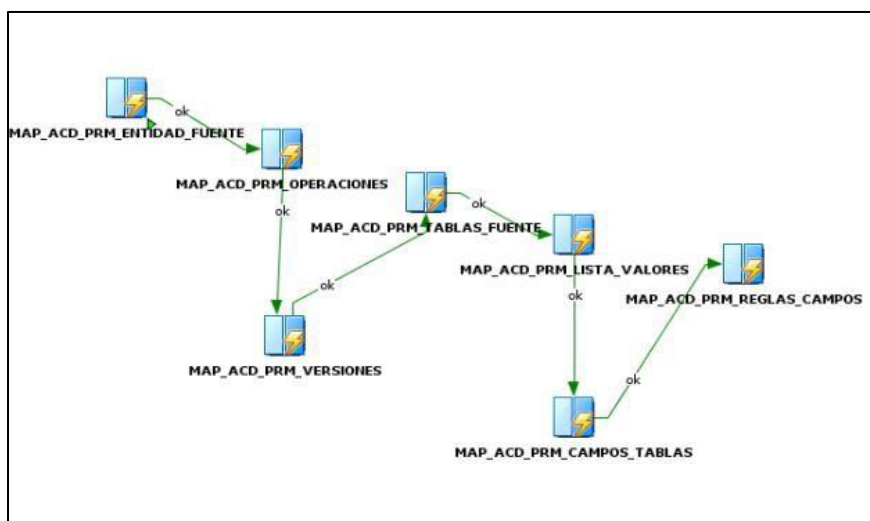
La lista de tablas que son cargadas por este JOB, son las siguientes:

- ACD_TRF_TIPO_DATOS
- ACD_TRF_CATEGORIAS_REGLAS
- ACD_TRF_TIPO_REGLAS
- ACD_TRF_REGLAS
- ACD_TRF_TIPO_EVALUACION
- ACD_TRF_TIPO_RESULTADO
- ACD_TRF_CRITERIOS_EVALUACION
- ACD_TRF_CLASIFICACIONES
- ACD_TRF_TIPO_CLASIFICACION

DESARROLLO

1.5.3. Cargue de datos de parámetros - PAQ_ACD_PRM_CARGAR

Gráfico 16. PAQ_ACD_PRM_CARGAR



Las tablas de parámetros se cargan a partir del PAQUETE de ODI, PAQ_ACD_PRM_CARGAR.

Consta de los pasos que permiten cargar cada una de las tablas de parámetros, tomando como fuente las hojas de del formato de especificaciones de Excel. Las tablas cargadas por cada uno de los pasos del Paquete, son las siguientes:

- ACD_PRM_ENTIDAD_FUENTE: Registra la información de la entidad o área responsable de la información a ser evaluada.
- ACD_PRM_OPERACION_FUENTE: registra la información referente a la temática a la cual pertenece la información a ser evaluada.
- ACD_PRM_VERSIONES: Versión en la que se actualiza traducción informática de reglas de validación.
- ACD_PRM_TABLAS_FUENTE: En esta hoja se registra los archivos a evaluar por cada base de datos.
- ACD_PRM_LISTA_VALORES:
- ACD_PRM_CAMPOS_TABLAS: En esta hoja se registra las variables o campos a evaluar por cada base de datos.
- ACD_PRM_REGLAS_CAMPOS: Identifica las reglas que se debe aplicar sobre cada uno de los campos a evaluar.

DESARROLLO

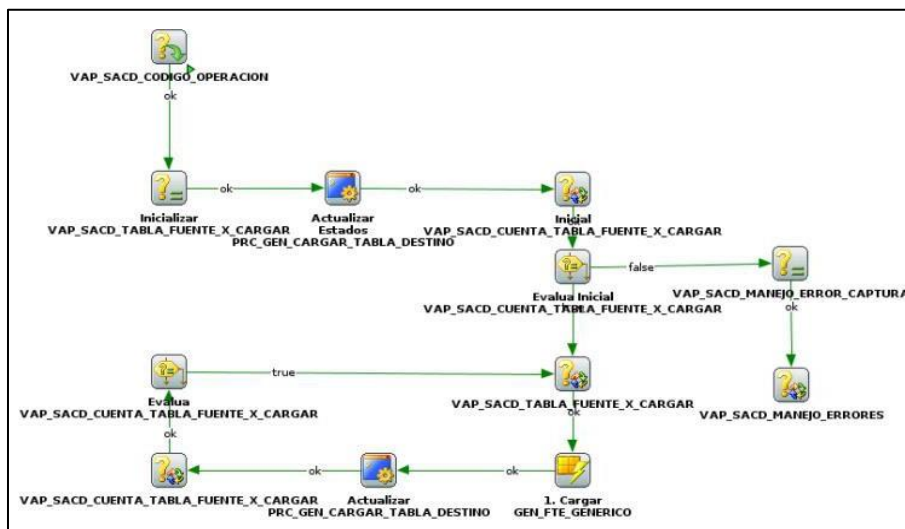
1.5.4. Cargue de archivos y tablas fuente

Este es un procedimiento genérico que permite identificar el tipo de fuente de cada una de las operaciones, es decir, se identifica si es un Excel XLSX, o un archivo plano independientemente de su extensión o separador de campos o si corresponde a una tabla de base de datos, para la cual se debe contar con la conexión de base de datos configurada en ODI, con su correspondiente Arquitectura Física y Lógica.

i) PAQ_GEN_CARGAR_TABLA_DESTINO

Este paquete realiza un ciclo de cargue a partir del nombre de la operación y ejecuta el proceso de cargue genéricos "PAQ_GEN_FUENTE_GENERICO", para cada una de las fuentes que se deben analizar, tomadas de la tabla "SACD_OAIEE.ACD_PRM_TABLAS_FUENTE", con el filtro sobre el campo "STR_CARGADA_X_FTE_GENERICO" = 'NO'; con el cual se cargan cada una de estas al área de trabajo, ubicada en el schema de base de datos "SACD_OAIEE".

Gráfico 17. PAQ_GEN_CARGAR_TABLA_DESTINO



Si el proceso no encuentra ninguna fuente para la operación a evaluar, se genera el error: 'NO SE ENCONTRARON FUENTES A EVALUAR PARA LA OPERACION ESDERSS' || 1/0

Solicita el siguiente parámetro:

- VAP_SACD_CODIGO_OPERACION: Sigla de la operación que se requiere ejecutar: Ejemplo NUSE, SIEDCO, MLCF_AV, etc.

DESARROLLO

○ **PAQ_GEN_FUENTE_GENERICO**

El paquete PAQ_GEN_FUENTE_GENERICO, realiza el cargue de un Excel o archivo plano de manera genérica sobre la tabla de trabajo “DAT_PLANO_GENERICO”, para luego de manera definitiva, dejar los datos en una tabla final por cada fuente de datos, sobre la que se realizan las validaciones necesarias para generar el indicador. La tabla final mantiene el prefijo DAT seguido del nombre de la tabla o archivo fuente a cargar; por ejemplo, el nombre de la tabla “DAT_PONALSIEDCOARMASXLSX” se conforma así:

DAT_ENTTOPPFUENTE

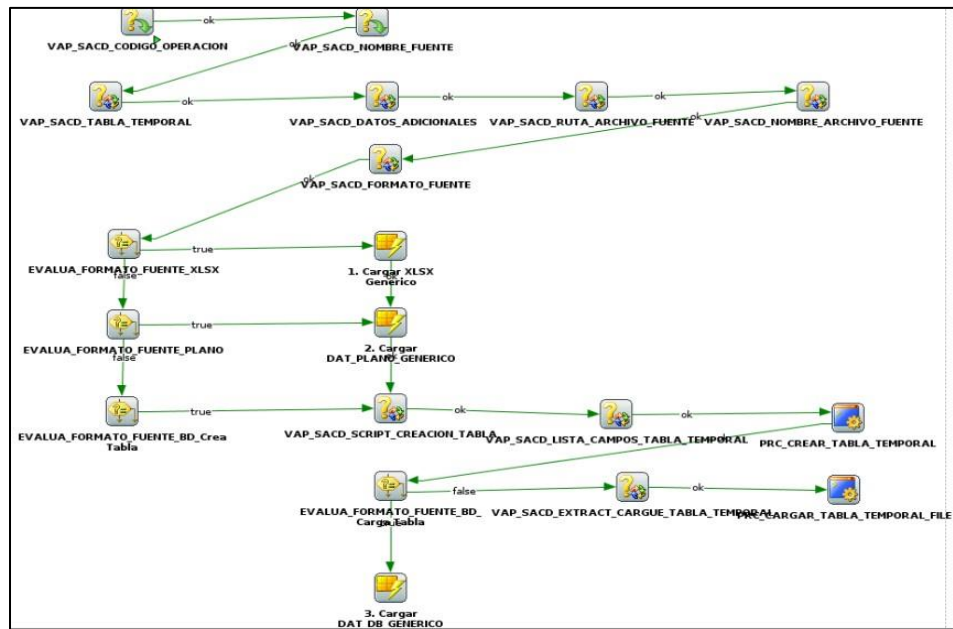
ENTT: Corresponde a la sigla de la entidad

fuelle. OPP: Sigla de la operación a aplicativo a

evaluar. FUENTE: Nombre de la tabla o archivo

fuelle.

Gráfico 18. PAQ_GEN_FUENTE_GENERICO



Solicita los siguientes parámetros:

- VAP_SACD_CODIGO_OPERACION: Sigla de la operación que se requiere ejecutar: Ejemplo NUSE, SIEDCO, MLCF_AV, etc.
- VAP_SACD_NOMBRE_FUENTE: Corresponde al nombre de la fuente que se

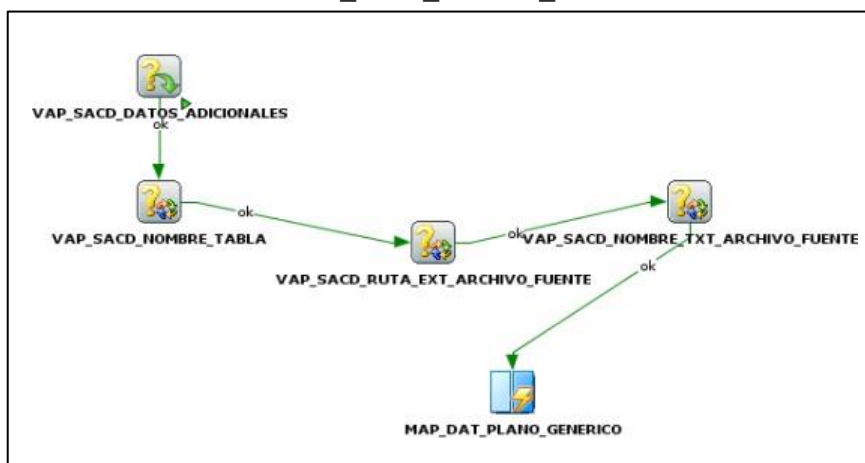
DESARROLLO

requiere cargar. Este parámetro cambio durante cada ejecución correspondiente a cada una de las fuentes de la operación.

○ PAQ_DAT_PLANO_GENERICO

Permite realizar el cargue de un archivo plano de cualquier tipo de separador, de la tabla de base de datos "DAT_PLANO_GENERICO", con una única columna, donde se concatenan cada uno de los campos del archivo.

Gráfico 19. PAQ_DAT_PLANO_GENERICO



Este paquete recibe el parámetro VAP_SACD_NOMBRE-TABLA; que corresponde al nombre definido de la tabla destino: Ejemplo "DAT_PONALSIEDCOARMASXLSX", el cual debe estar registrado en la tabla de control-carga, junto con la ruta o ubicación del archivo en el servidor de archivo. Este nombre va a servir para refrescar las siguientes dos variables que son la ruta del archivo y el nombre del archivo, tomándolos de la tabla de control.

- VAP_SACD_RUTA_EXT_ARCHIVO_FUENTE: Es la ruta del servidor ODI donde se encuentra ubicado el archivo fuente. Ejemplo:
/u01/Dataserver/SACD/
- VAP_SACD_NOMBRE_TXT_ARCHIVO_FUENTE: El nombre del archivo fuente. Ejemplo: Armas.txt

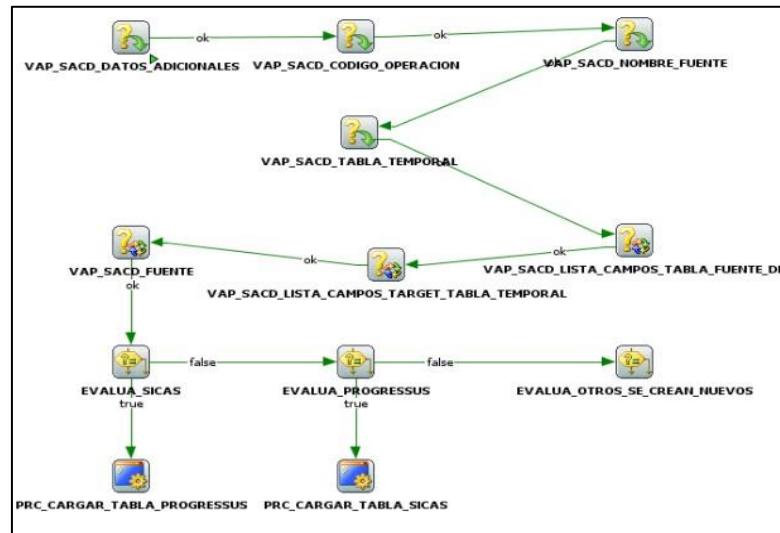
○ PAQ_DAT_BD_GENERICO

Hace parte del paquete principal PAQ_GEN_FUENTE_GENERICO. Su funcionalidad es realizar el cargue de una tabla de la base de datos fuente a una tabla en el ambiente del SACD, es decir en el schema "SACD_OAIEE".

DESARROLLO

Inicialmente toma como referencia cada uno de los campos definidos en el formato de entrada para una tabla específica, y los consulta en la tabla fuente para luego crear una tabla destino y cárgala en la tabla final del schema "SACD_OAIEE", con cada uno de los campos y registros consultados.

Gráfico 20. PAQ_DAT_BD_GENERICO



Este paquete puede requerir modificaciones adicionales en la medida en que existan nuevas fuentes de datos, para lo cual se debe crear la arquitectura lógica y física de ODI, al igual que el procedimiento que realiza la lectura de la tabla fuente y la inserción en el destino, el cual se debe definir para cada una de las fuentes que se requiera evaluar. Es este caso lo se tiene las fuentes de PROGRESSUS y SICAS.

1.5.5. Ejecución general del cálculo

Efectúa de manera ordenada las invocaciones de los procesos para el cargue de calidad de datos. Esto estos pasos están incluidos en el paquete de base de datos "PKG_SACD_CALCULAR_ACD".

i) Descripción de los procedimientos:

- **PRC_CALCULAR_ACD**

Procedimiento orquesta los procesos de cálculo del Aseguramiento de Calidad de Datos; aplicación de reglas, cálculo de indicadores y obtención de resultados.

DESARROLLO

PARAMETROS :

IN:

P_OPERACION nombre de la operación a evaluar

- ✓ aplicación de reglas

PKG_ACD_RSL_REGISTROS_INVALIDS.PRC_EJECUTAR_REGLAS(P_OPERACION => P_OPERACION);

- ✓ cálculo de indicadores

PKG_ACD_RSL_DETALLE_RESULTADOS.PRC_CALCULAR_INDICADOR_CALIDAD (P_OPERACION => P_OPERACION);

- ✓ obtención de resultados

PKG_ACD_RSL_RESUMEN_RESULTADOS.PRC_CARGAR_RESUMEN_RESULTADOS(P_OPERACION => P_OPERACION);

Este es el proceso que realiza el cálculo de los indicadores a partir de todos los registros de las tablas o archivos fuente; identificando total de registros de cada una, registros defectuosos por tabla y generando los indicadores de calidad.

- **Identificación de registros Inválidos.**

PAQUETE: PKG_ACD_RSL_REGISTROS_INVALIDS

PROCEDIMIENTO: PRC_EJECUTAR_REGLAS (P_OPERACION)

Este es el proceso que realiza la aplicación de reglas de validación sobre todos los registros de las tablas o archivos fuente, con el fin de identificar registros inválidos y tomarlos en cuenta al momento de calcular los indicadores.

Recibe como parámetro la operación "P_OPERACION", con la que se realiza el procesamiento de todas las fuentes de cada operación y las reglas configuradas por cada una de ellas.

Obtiene cada una de las reglas parametrizadas en la hoja "4.IdentificaReglas", del formulario de especificaciones, y aquellas que se identifican de manera genérica o que no requieren traducción informática. Cada una de estas reglas es aplicada sobre el campo correspondiente y se identifican los registros que no cumplen el criterio para ser enviados a la tabla "ACD_RSL_REGISTROS_INVALIDOS". En esta tabla se envía el registro con su identificador único y todo su contenido en el campo "CLB_CONTENIDO_REGISTRO", así como el query con el que se realiza la validación del registro.

DESARROLLO

ii) Cálculos Generales del indicador.

Una vez completada la identificación de registros rechazados, se inicia el proceso que permite calcular los diferentes indicadores, a partir del conteo de registros totales por tablas, registros inválidos, número de columnas o campos, campos afectados, etc. Dichos procesos están incluidos en las funciones del paquete de base de datos "PKG_ACD_CALCULOS_GENERALES".

Descripción de los procedimientos:

1) *fnc_calcular_registros_defectuosos*

Función que recibe como parámetro el nombre de una tabla y la operación y devuelve el número de registros defectuosos; entendiéndose por registro cada una de las filas de una tabla.

PARAMETROS:

IN:

P_OPERACION nombre de la operación a evaluar P_TABLA nombre de la tabla a evaluar

OUT:

V_NUM_REGISTROS: Retorna el número de registros defectuosos

2) *fnc_calcular_total_registros_fuente*

Esta función recibe como parámetro el nombre de una tabla y devuelve el número de registros cargados en la tabla fuente; entendiéndose por registro cada una de las filas de la tabla.

PARAMETROS:

IN:

P_TABLA nombre de la tabla a evaluar

OUT:

V_NUM_REGISTROS: Retorna el número de registros

DESARROLLO

3) fnc_calcular_campos_con_defectos

Función que recibe como parámetro el nombre de una tabla y la operación y devuelve el número de campo defectuosos; entendiéndose por campo cada una de las intercepciones entre fila y una columna de una tabla.

PARAMETROS:

IN:

P_OPERACION nombre de la
operación a evaluar P_TABLA
nombre de la tabla a evaluar

OUT:

V_NUM_REGISTROS: Retorna el total de campos con defectos

4) fnc_calcular_total_campos_fuente

Función que recibe como parámetro el nombre de una tabla y devuelve el número de campos; entendiéndose por campo cada una de las intercepciones entre fila y una columna de la tabla consultada.

PARAMETROS:

IN:

P_TABLA nombre de la tabla a evaluar

OUT:

v_campos Retorna el número de campos

5) fnc_calcular_variables_con_defectos

Función que recibe como parámetro el nombre de una tabla y la operación y devuelve el número de variables con defectos; entendiéndose por variables cada una de las columnas de la tabla que presentaron algún defecto.

PARAMETROS:

IN:

P_TABLA nombre de la tabla a evaluar

OUT:

V_NUM_REGISTROS: Retorna el total de variables con defectos

DESARROLLO

6) fnc_calcular_total_variables_campos_fuente

Función que recibe como parámetro el nombre de una tabla y devuelve el número total de variables de la fuente; entendiéndose por variables cada una de las columnas de la tabla fuente.

PARAMETROS:

IN:

P_TABLA nombre de la tabla a evaluar

OUT:

V_NUM_REGISTROS: Retorna total de variables

7) fnc_obtener_ultima_version

Función que recibe como parámetro el código de una operación y devuelve la llave subrogada de la última versión cargada por código de la versión.

PARAMETROS :

IN:

P_OPERACION nombre de la operación a evaluar

OUT:

V_VERSION Retorna la sk de la versión

8) fnc_calcular_registros_x_regla

Función que recibe como parámetro el nombre de una tabla y la operación y devuelve el número de registros defectuosos por categoría de la regla; entendiéndose por registro cada una de las filas de una tabla.

PARAMETROS:

IN:

P_OPERACION nombre de la
operación a evaluar P_TABLA

nombre de la tabla a evaluar

P_TIPO_REGLA: Puede ser cualquiera de los siguientes:

DESARROLLO

IDENTIDAD, INTEGRIDAD_REFERENCIAL, CARDINALIDAD, HERENCIA, RELACIONES_DEPENDIENTES, DOMINIO ATRIBUTOS_DEPENDIENTES, DEFINICION, CUBRIMIENTO_DE_VALORES, DUPLICACION, CORRECTITUD, DISCREPANCIA OUT:

OUT:

V_NUM_REGISTROS Retorna el número de registros inválidos por regla

9) fnc_calcular_variables_analizados

Función que recibe como parámetro el nombre de una tabla y la operación y devuelve el número de variables analizadas; entendiéndose por variables cada una de las columnas de una tabla.

PARAMETROS:

IN:

P_OPERACION nombre de la operación a evaluar P_TABLA nombre de la tabla a evaluar

OUT:

V_NUM_REGISTROS Retorna el número de variables analizadas por tabla

10) fnc_calcular_variables_con_defectosVar

Función que recibe como parámetro el nombre de una tabla, campo y la operación y devuelve el número de registros por variable con defectos; entendiéndose por variables cada una de las columnas de la tabla que presentaron algún defecto.

PARAMETROS:

IN:

P_OPERACION nombre de la operación a evaluar P_TABLA nombre de la tabla a evaluar P_CAMPO nombre del campo a evaluar

OUT:

V_NUM_REGISTROS Retorna el total de variables con defectos

DESARROLLO

11) fnc_calcular_campos_analizados

Función que recibe como parámetro el nombre de una tabla y la operación y devuelve el número de campos analizados; entendiéndose por campo cada una de las intercepciones entre fila y una columna de la tabla.

PARAMETROS:

IN:

P_OPERACION nombre de la
operación a evaluar P_TABLA
nombre de la tabla a evaluar

OUT:

V_NUM_REGISTROS Retorna el número de campos analizados por tabla

12) fnc_calcular_archivos_analizados

Función que recibe como parámetro el nombre de una operación y devuelve el número de tablas analizadas.

PARAMETROS :

IN:

P_OPERACION nombre de la operación a evaluar

OUT:

V_NUM_REGISTROS Retorna el número de tablas analizadas

13) fnc_calcular_registros_analizados

Función que recibe como parámetro el nombre de una tabla y devuelve el número de campos; entendiéndose por campo cada una de las intercepciones entre fila y una columna de la tabla.

PARAMETROS:

IN:

P_OPERACION nombre de la
operación a evaluar P_TABLA

DESARROLLO

nombre de la tabla a evaluar

OUT:

V_NUM_REGISTROS Retorna el número de campos por tabla

14) fnc_calcular_reglas_aplicadas

Función que recibe como parámetro el nombre de una tabla y devuelve el número de reglas aplicadas.

PARAMETROS:

IN:

P_OPERACION nombre de la
operación a evaluar P_TABLA
nombre de la tabla a evaluar

OUT:

V_NUM_REGISTROS Retorna el número de reglas por tabla

iii) Evaluación detallada.

Referencia cada una de las funciones de cálculos generales, obtener resultados en la tabla finales del proceso de aseguramiento de calidad, a nivel de registros, campos y variables. Dichos procesos están incluidos en las funciones del paquete de base de datos "PKG_ACD_RSL_DETALLE_RESULTADOS"

Descripción de los procedimientos:

1) PRC_CARGAR_EVALUACION_DETALLADA

Procedimiento que obtiene el detalle de los indicadores por registro, campo y variable y los carga a la tabla ACD_RSL_DETALLE_RESULTADOS. El detalle de indicadores a almacenar es: Registro_total, registro_invalidos, registro_indefinidos, registro_faltantes, registro_duplicados, registro_incorrectos, registro_discrepantes, registro_inconsistentes, campo_inconsistentes, campo_total, variable_inconsistentes, variable_analizados, variable_total.

DESARROLLO

PARAMETROS:

IN:

P_OPERACION nombre de la operación a evaluar

2) PRC_CALCULAR_INDICADOR_CALIDAD

Procedimiento que calcula el indicador de calidad ponderado para los tipos de evaluación

'REGISTRO_INCONSISTENTES','CAMPO_INCONSISTENTES','VARIABLE_INCONSISTENTES' y los carga a la tabla ACD_RSL_DETALLE_RESULTADOS.

PARAMETROS:

IN:

P_OPERACION nombre de la operación a evaluar

○ **Resumen de resultados.**

Se toman como referencia la evaluación detallada a nivel de registro, variable y campo "PKG_ACD_RSL_RESUMEN_RESULTADOS", para calcular los indicadores finales de calidad y total.

Descripción de los procedimientos:

3) PRC_CARGAR_RESUMEN_RESULTADOS

Procedimiento que obtiene el resumen de los indicadores por registro y los carga a la tabla ACD_RSL_RESUMEN_RESULTADOS.

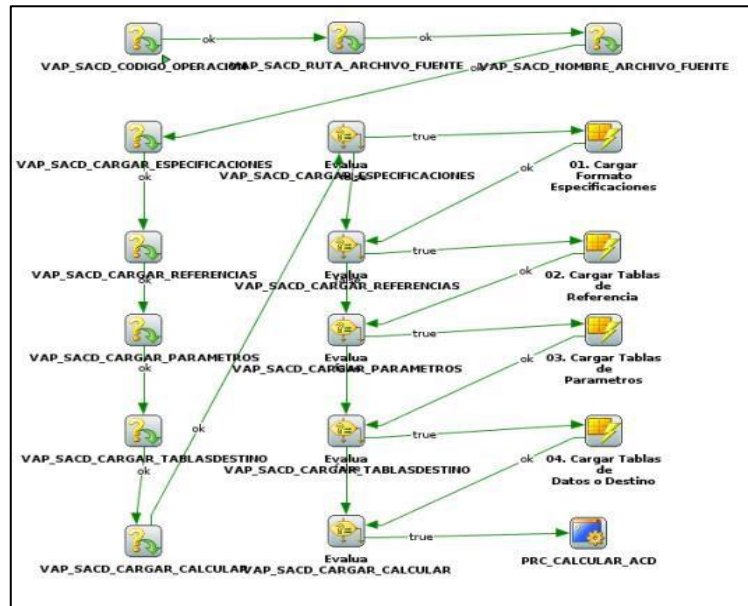
PARAMETROS:

IN: P_OPERACION nombre de la operación a evaluar

1.5.6. PAQ_ACD_PROCESAMIENTO

Corresponde al paquete principal que orquesta todo el proceso para el aseguramiento de calidad de datos; desde el cargue de especificaciones hasta el procesamiento de datos y cargue de tablas de resultados del ACD.

Gráfico 21. PAQ_ACD_PROCESAMIENTO



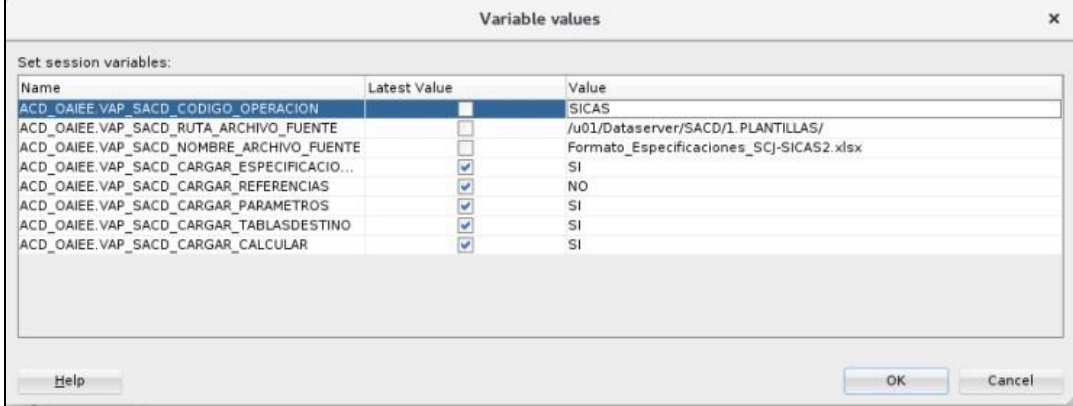
i) Parámetros de Entrada

Lista de parámetros de entrada que requiere el paquete, para la ejecución, cargue y procesamiento del indicador.

Dependiendo del valor indicado en los parámetros; se pueden realizar u omitir algunos pasos en el proceso de cargue:

DESARROLLO

Gráfico 22. Parámetros de entrada



Name	Latest Value	Value
ACD_OAIEE.VAP_SACD_CODIGO_OPERACION	<input type="checkbox"/>	SICAS
ACD_OAIEE.VAP_SACD_RUTA_ARCHIVO_FUENTE	<input type="checkbox"/>	/u01/Dataserver/SACD/1.PLANTILLAS/
ACD_OAIEE.VAP_SACD_NOMBRE_ARCHIVO_FUENTE	<input type="checkbox"/>	Formato_Especificaciones_SCJ-SICAS2.xlsx
ACD_OAIEE.VAP_SACD_CARGAR_ESPECIFICACIONES...	<input checked="" type="checkbox"/>	SI
ACD_OAIEE.VAP_SACD_CARGAR_REFERENCIAS	<input checked="" type="checkbox"/>	NO
ACD_OAIEE.VAP_SACD_CARGAR_PARAMETROS	<input checked="" type="checkbox"/>	SI
ACD_OAIEE.VAP_SACD_CARGAR_TABLASDESTINO	<input checked="" type="checkbox"/>	SI
ACD_OAIEE.VAP_SACD_CARGAR_CALCULAR	<input checked="" type="checkbox"/>	SI

Descripción de Variables de entrada del escenario:

- ACD_OAIEE.VAP_SACD_CODIGO_OPERACION: corresponde al código de operación que se va a cargar.
- ACD_OAIEE.VAP_SACD_RUTA_ARCHIVO_FUENTE: Es la ruta del servidor ODI donde se encuentra ubicado el archivo.
Ejemplo:
/u01/Dataserver/SACD/1.PLANTILLAS/
- ACD_OAIEE.VAP_SACD_NOMBRE_ARCHIVO_FUENTE: El nombre del formato de entrada de datos. Ejemplo: Formato_Especificaciones_SCJ-SIEDCO.xlsx
- ACD_OAIEE.VAP_SACD_CARGAR_ESPECIFICACIONES: Identifica SI se cargará o NO el formato de especificaciones definido con el nombre de archivo y ruta fuente definidos en las variables anteriores.
- ACD_OAIEE.VAP_SACD_CARGAR_REFERENCIAS: Indica que se van a cargar las tablas de referencia. Esto solo se hace una sola vez; y no es necesario volver a realizarlo.
- ACD_OAIEE.VAP_SACD_CARGAR_PARAMETROS: Indica que se van a cargar las tablas de parámetros, al sistema relacional; es decir los datos del formato de especificaciones se cargan al sistema relacional, generando llaves únicas, para cada uno de las variables y tablas definidas en el formato.
- ACD_OAIEE.VAP_SACD_CARGAR_TABLASDESTINO: define SI se cargaran cada una de las tablas del sistema fuente a su correspondiente tabal DAT_ENTTOPPFUENTE
- ACD_OAIEE.VAP_SACD_CARGAR_CALCULAR: Indica que se va a realizar el cálculo del indicador con las nuevas reglas que se carguen del formato de especificaciones.

DESARROLLO

- ACD_OAIEE.VAP_SACD_CARGAR_RESULTADOS: Indica que se deben generar los archivos con las salidas de información, del indicador y los registros inválidos.

1.5.7. Parametrización cargar tabla

Cada vez se defina una nueva tabla fuente; es necesario incluir su nombre en la tabla de control. Para esto se debe obtener el nombre de la tabla indicado en el campo "ACD_PRM_TABLAS_FUENTE.STR_NOM_TABLA_TEMPORAL". Con este nombre de tabla es que realizan todas las validaciones referentes a la calidad del dato.

El nombre temporal está dado por las iniciales de la entidad, seguido de la sigla de la operación, y el nombre de la tabla sin espacios ni caracteres especiales.

Por ejemplo, para el caso de SICAS, se definieron los siguientes nombres de tablas temporales:

Gráfico 23. Nombres de tablas temporales

STR_NOM_TABLA_FUENTE_NK	STR_NOM_TABLA_TEMPORAL
CASA JUSTICIA	DAT_SCJSICASCASAJUSTICIA
LOCALIDAD	DAT_SCJSICASLOCALIDAD
NACIONALIDAD	DAT_SCJSICASNACIONALIDAD
PERMISO_SICAS	DAT_SCJSICASPERMISOSICAS
ESTRUCTURA_FAMILIAR	DAT_SCJSICASESTRUCTURAFAMILIAR
COMO_SE_ENTERO	DAT_SCJSICASCOMOSEENTERO
BARRIO	DAT_SCJSICASBARRIO
HABITANTE	DAT_SCJSICASHABITANTE
PROBLEMA_CRI	DAT_SCJSICASPROBLEMACRI
TIPO_DOCUMENTO	DAT_SCJSICASTIPODOCUMENTO
CARACTERIZACION_CRI	DAT_SCJSICASCARACTERIZACIONCRI
PROBLEMA_CARACTER...	DAT_SCJSICASPROBLEMACARACTERIZ
VISITA_CASA JUSTICIA	DAT_SCJSICASVISITACASAJUSTICIA
CIUDAD	DAT_SCJSICASCIUDAD
DEPARTAMENTO	DAT_SCJSICASDEPARTAMENTO
ENTERO	DAT_SCJSICASENTERO
FUNCIONARIO_CASA	DAT_SCJSICASFUNCIONARIOCASA
DIRIGIDO_DESDE_CRI	DAT_SCJSICASDIRIGIDODESDECRI

Elaboró: Juan Carlos Bulla Abril – Contratista Oficina de Análisis de Información y Estudios Estratégicos

Revisó: Diana Marcela Flechas Ruiz - Contratista Oficina de Análisis de Información y Estudios Estratégicos

La información de aprobación de este documento podrá ser consultada en el sistema “Portal MIPG” - <https://portalmipg.scj.gov.co>