

INFORME DE DESARROLLO

TAREAS DESARROLLADAS PARA ETL DE INGESTA UIAF

Jorge Andrés Salazar

08/03/2021

CONTENIDO

INTRODUCCIÓN	3
OBJETIVO	3
DESCRIPCIÓN DE LAS ACTIVIDADES.....	3
1. Actividades	3
2. Cronograma de actividades.....	6

INTRODUCCIÓN

A continuación, se detallará todas las actividades realizadas en el cluster de desarrollo de UIAF, con sus respectivas fechas de tareas efectuadas, las tareas son aplicadas a distintos procesos ETL y su desarrollo se realizó para atender los requerimientos de gestión de carga y su respectiva generalización.

OBJETIVO

Desarrollar procesos ETL para ingesta en el cluster de UIAF.

DESCRIPCIÓN DE LAS ACTIVIDADES

En el transcurso del desarrollo de las actividades el seguimiento se realizó mediante la herramienta Trello; se explicará los detalles de las actividades realizadas a continuación:

1. ACTIVIDADES

1. Verificar modelo estrella efectivo.

- Generación de archivos de texto plano para Efectivo y Productos generales a partir del anexo técnico.

2. Automatizar ADRES como paso 2.

- Función de postgres para obtención de procedimiento de paso 2 get_procepasso_2 con registro de tablas origen-destino para carga de paso 2.
- Se escribió el etl_ADRES.py para carga de paso 2 de las tablas dim. Funciones escritas:
 - Ingesta_paso_2
 - carga_origen
 - cargar_tabla
 - procesamiento_right_sistema
 - procesamiento_inner

El ETL de ADRES se escribió conforme a los requerimientos de gestión de carga en la base de datos de postgres.

3. PG110 y PG210 pruebas con idcargue.

- Se confirmo con los PG110 y PG220 que el ETL_efectivo funciona correctamente. El ETL_efectivo se escribió conforme a los requerimientos de la base de datos de gestión.

4. Insert de Productos Generales en postgres.

- Script submodelo_9_Jorge.sql para carga de submodelo uiaf_productos a base de datos de gestión.

5. La migración a dwh_uiaf_2.

- Ajustes y migración a dwh_uiaf_2 en hive por medio del script adaptacion_a_uiaf_2.py subido en repositorio de gitlab.
- Modificaciones a la base de datos dwh_uiaf por medio del script de hiveql reformas dwh_uiaf_2.sql.

6. Cambio de nombre de las tablas de modelo a detallado.

- Reformas en base de datos dwh_uiaf2, cargadas en sql al repositorio de gitlab.

7. Documentar Funciones de ADRES y Efectivo.

- Carga de uiaf_modelo.pdm modelo entidad-relación a Gitlab.
- Se realizo la documentación de los siguientes scripts:
- -preprocesamiento.py
- -tareas_1.py
- -tareas_2.py
- -etl_ADRES.py
- -etl_Efectivo.py
- -common.py (nueva funcionalidad)

8. Desarrollar Paso 2 de Efectivo y PG110 y PG210.

- Scripts, efectivo paso 2, preprocesamiento, cruce inner cargados al gitlab.
- Inserts configurados y subidos al Gitlab.

9. Ajuste de funciones pyspark con el nuevo esquema de uiaf_modelo.PDB.

- Se ha trabajado Efectivo con una configuración por dimensiones, con los siguientes scripts escritos:
- cruze_right.py
- cruze_inner.py
- ETL_general.py
- inner_dimdane.py
- inner_dimmoneda.py
- inner_dimpersona.py
- inner_dimproducto.py
- inner_dimtipoidentificacion.py
- inner_dimtipoproducto.py
- inner_dimactividad.py
- inner_dimpaises.py
- inner_factefectivo (Si es necesario)
- preproc_dimdane.py
- preproc_dimmoneda.py
- preproc_dimpersona.py
- preproc_dimproducto.py
- preproc_dimtipoidentificacion.py
- preproc_dimtipoproducto.py
- preproc_factefectivo.py
- preproc_dimactividad.py
- preproc_dimpaises.py
- right_dimdane.py

- right_dimmoneda.py
- right_dimpersona.py
- right_dimproducto.py
- right_dimtipoidentificacion.py
- right_dimtipoproducto.py
- right_dimactividad.py
- right_dimpaises.py
- Con esta configuración se puede generalizar respectivamente para recibir los demás sistemas a los cuales se les quiera realizar la ingesta, ya sea ADRES o Productos generales, para la respectiva ejecución de un solo script de carga.

2. CRONOGRAMA DE ACTIVIDADES

A continuación, se presenta el siguiente orden cronológico de las actividades que se han efectuado.

Tabla 1. Cronograma de actividades.

Actividad	Fecha inicio	Fecha Finalizado
Verificación del modelo estrella efectivo y productos generales.	25 de nov de 2020	27 de nov de 2020
Automatizar ADRES como paso 2.	1 de dic de 2020	7 de dic de 2020
PG110 y PG210 pruebas con idcargue.	4 de dic de 2020	7 de dic de 2020
Desarrollar paso 2 de efectivo y PG110 y PG210 con nueva lógica por dimensiones. (Se continuaría con el paso 2 de productos generales)	7 de dic de 2020	5 de marzo de 2021
Insert de productos generales en postgres.	7 de dic de 2020	12 de feb de 2021
Documentar funciones de ADRES y Efectivo	7 de dic de 2020	8 de feb de 2021
Documentación de los scripts preprocesamiento, tareas_1, tareas_2, etl_ADRES.py, etl_Efectivo.py, common.py	10 de dic de 2020	18 de dic de 2021
Cambio de nombre de las tablas de modelo a detallado (Se llevo a cabo por medio de 2 flujos de trabajo)	12 de feb de 2021	25 de feb de 2021
Ajuste de funciones pyspark con el nuevo esquema de uiaf_modelo.PDB	12 de feb de 2021	25 de feb de 2021
Migración a dwh_uiaf_2	12 de feb de 2021	25 de feb de 2021
Documentación de la nueva lógica por dimensiones aplicada.	3 de marzo de 2021	8 de marzo de 2021