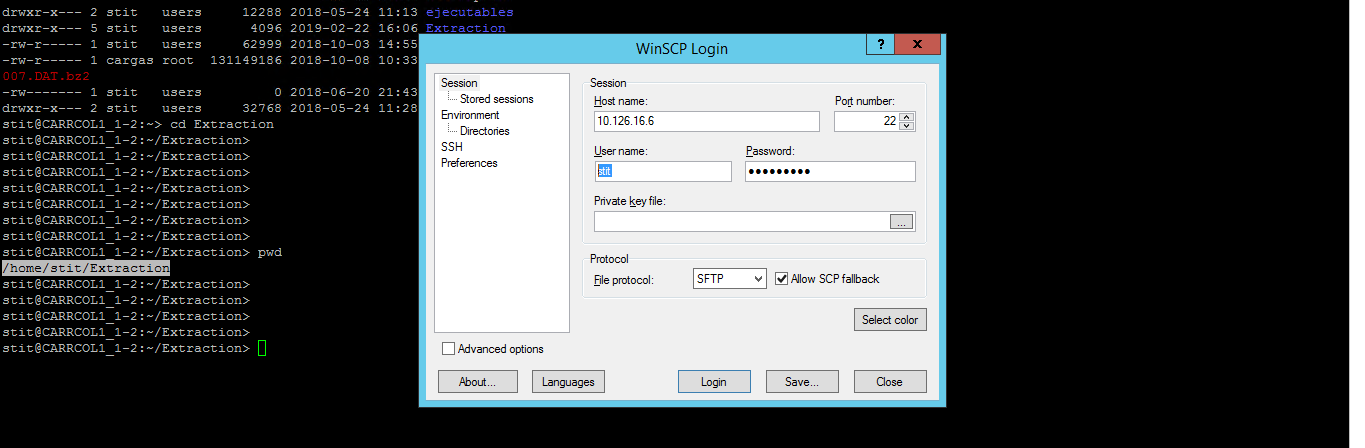
**DEFINICIÓN DE EXTRACCIÓN**

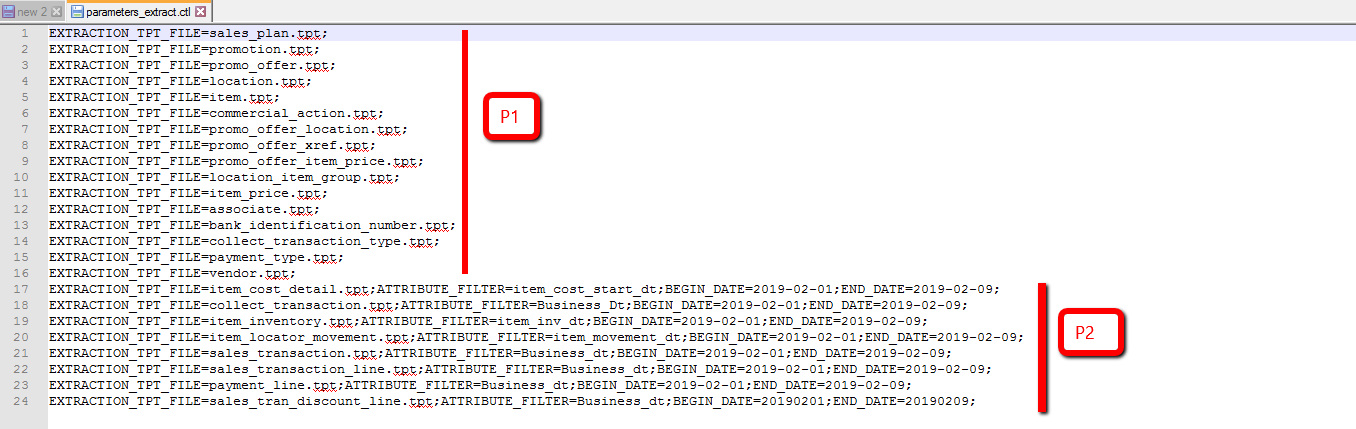
Login in Host: “10.126.16.6”

User: “stit”

Password: “c3nc0sud.”



En el directorio "/ home / stit / Extraction" editar el archivo "parameters\_extract.ctl"



La sesión "P1" son las tablas que se van a realizar extracción total de la tabla, por lo que estas líneas no se deben cambiar

La sesión "P2" son las tablas que se van a realizar extracción de un período, donde cada línea es una extracción diferente, se deben cambiar las fechas del período de extracción:

- EXTRACTION\_TPT\_FILE, es el archivo que va a hacer la extracción, donde contiene el nombre de la tabla con sufijo .tpt (ejemplo: sales\_transaction.tpt)

- ATTRIBUTE\_FILTER, nombre de la columna que define el período de extracción (ejemplo: Business\_dt)

- BEGIN\_DATE, inicio del período de extracción (ejemplo: 2019-02-01)

- END\_DATE, fin del período de extracción (ejemplo: 2019-02-07)

Atención: cada línea de extracción crea un archivo; para no generar un error de espacio de memoria (cuando el archivo es grande), cada línea debe contener un máximo de 15 días;

Si el período de extracción es mayor que 15 días, se debe crear otra línea de extracción para la misma tabla.

ejemplo:

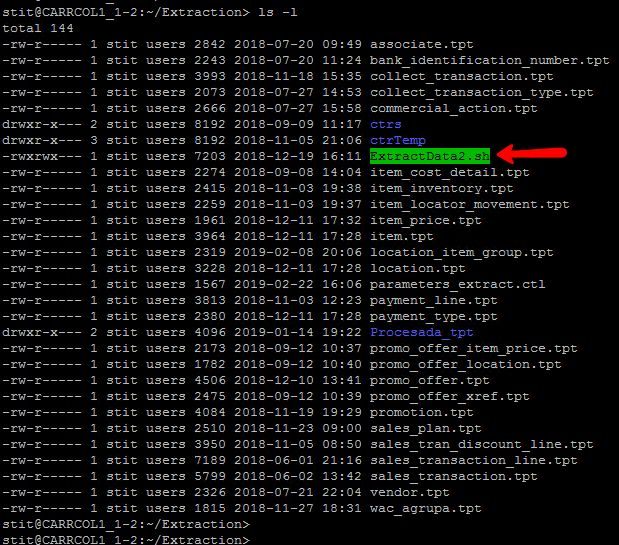
40 días

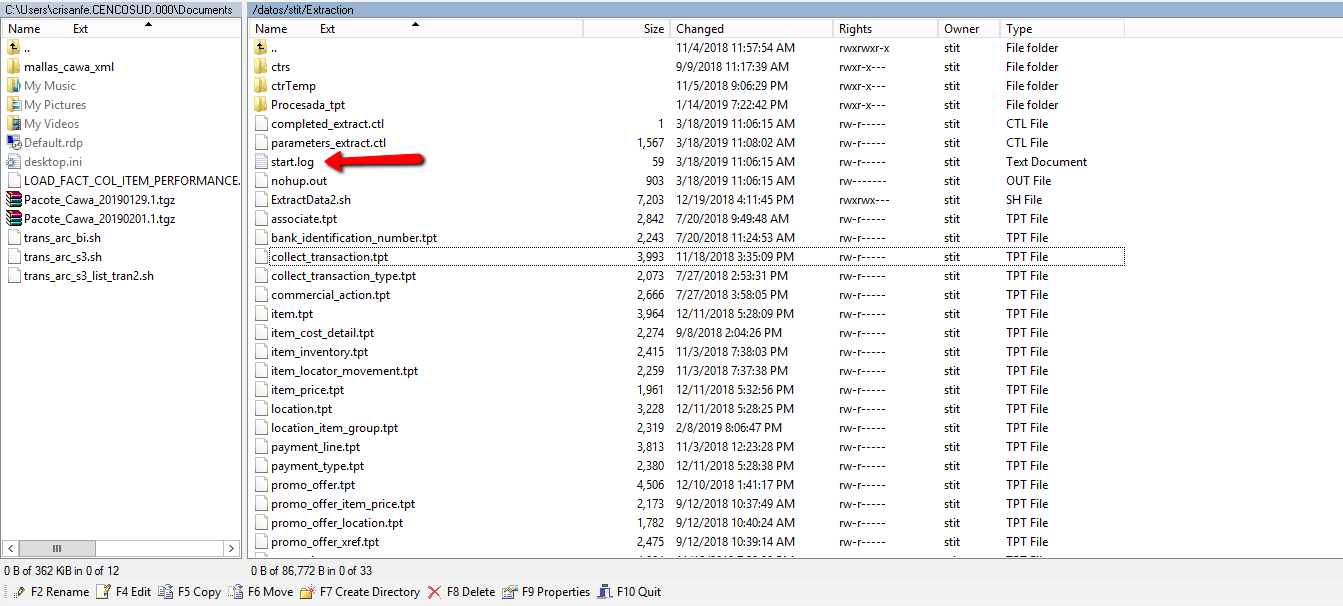
EXTRACTION\_TPT\_FILE=sales\_transaction.tpt;ATTRIBUTE\_FILTER=Business\_dt;BEGIN\_DATE=2019-01-01;END\_DATE=2019-01-15; EXTRACTION\_TPT\_FILE=sales\_transaction.tpt;ATTRIBUTE\_FILTER=Business\_dt;BEGIN\_DATE=2019-01-16;END\_DATE=2019-01-31; EXTRACTION\_TPT\_FILE=sales\_transaction.tpt;ATTRIBUTE\_FILTER=Business\_dt;BEGIN\_DATE=2019-02-01;END\_DATE=2019-02-10;

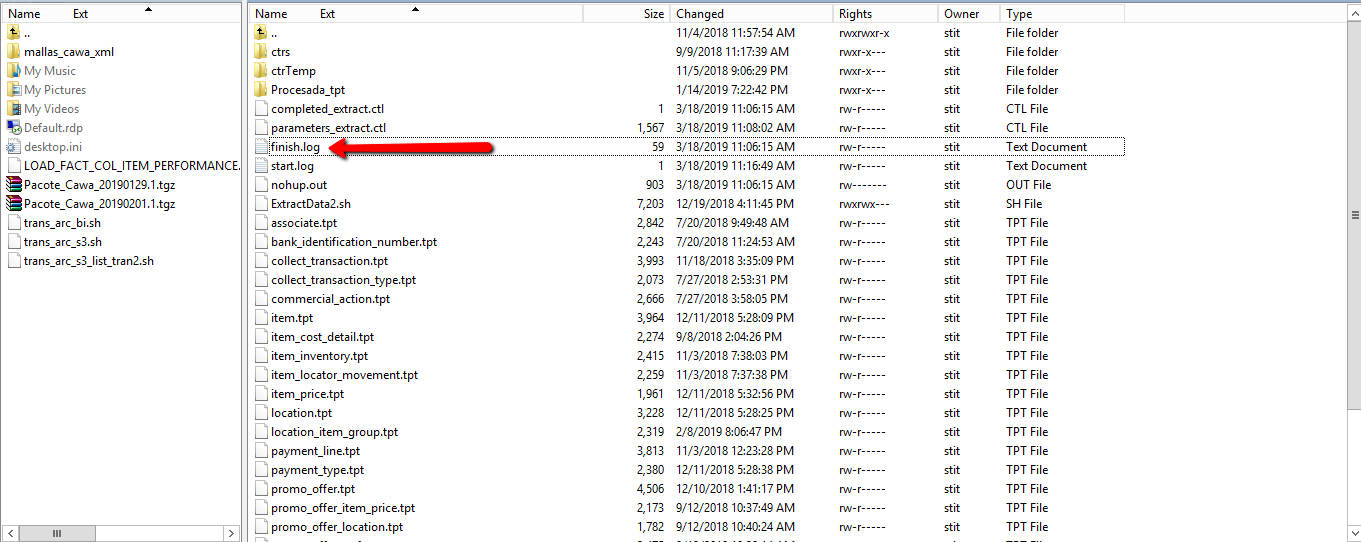
**Ejecutar script de extracción**

Después de editar y save el archivo "parameters\_extract.ctl" por pantalla de comandos se debe ejecutar el script "ExtractData2.sh", con el siguiente comando:

nohup ./ExtractData2.sh stit.migracion c3nc0sud dwh\_data &



Esta ejecución puede tardar, y los logs se ejecutan en backgroud, por lo que debe observarse, cuando el job de extracción está en ejecución, va a crear el archivo "start.log"

Cuando el job termina la extracción, se va a crear un archivo "finish.log", y si no hay ningún mensaje de error escrito en el archivo, la extracción se realizó correctamente.



**Cargando el datalake**

Después de finalizar la extracción ejecutar el job glue "job\_extracion" para cargar el datalake

