

DOCUMENTACION TP1 GRUPO03

Alfonso Javier Amezcua - 94732

1) Diagrama de la estructura antes de la instalación:

Una vez descargado el directorio GRUPO03 se podrá ver que dentro del mismo hay 5 subdirectorios. Una breve descripción de cada uno y de su contenido es la siguiente:

- **inst:** Es el directorio de instalación, dentro esta el script instalador.
- **scripts:** dentro estan los scripts de start (arranque del proceso), stop (detención del proceso), inicializador y proceso.
- **docs:** dentro hay:
 - un archivo log de instalación que tiene registrado distintos escenarios como sistema ya instalado, sistema instalado pero necesita reparacion, reparación del sistema y reinstalación del mismo.
 - un archivo de configuración resultante luego de una instalación exitosa.
 - un archivo de log de inicialización que tiene distintos escenarios como inicialización del sistema, indicación que un sistema no se puede inicializar por instalación rota, inicialización exitosa y no se puede inicializar por ya inicializado.
 - un archivo de log de proceso con descripción para la ejecución de un proceso mostrando cómo el proceso avanza ciclo a ciclo y procesa todos los casos de prueba que están en el directorio pruebas.
- **tablas_cpr:** contiene las tablas de comercio, de provincias y de respuestas del gateway. Todas en formato .csv.
- **pruebas:** dentro de este directorio hay subdirectorios que contienen pruebas que se usaron para probar el funcionamiento del proceso. Los subdirectorios están nombrados según el tipo de prueba que guardan:
 - **aceptados:** son los archivos que nos entregaron con el enunciado, todos estos son archivos cuyos registros no son rechazados y generan archivos de salida válidos.

- **rechazados:** son archivos que serán enviados al directorio de rechazados ya sea porque su formato de nombre es inválido, porque está vacío, por no ser .txt o porque ya se proceso un archivo con nombre similar. **Dentro de cada uno de ellos hay una descripción de porque fue rechazado.**
- **fin de mes:** acá hay 12 archivos que en su nombre tienen el ultimo día de cada mes, sirven para demostrar que en la salida el archivo resultante de cada uno tiene el nombre del primer día del siguiente mes.
- **aceptados_err_campos:** estos archivos serán aceptados y mandados al directorio de aceptados pero pueden tener en su contenido un registro que será rechazado y mandado al archivo RejectedData.
- **ajustar fecha:** solo tiene un archivo, si el sistema tiene una hora de cierre menor a 180000 entonces la fecha del nombre va a avanzar un día en el archivo de salida, si no la tiene entonces no. Solo sirve para mostrar que se logra este hecho.

Diagrama de la estructura después de la instalación:

Si la instala exitosamente el sistema entonces el directorio GRUPO03 resultante tiene la siguiente forma:

- **Directorio inst:** contiene:
 - instalador.sh
 - instalador.log
 - instalador.conf
- **Directorio de tablas (tab por defecto):** guarda las tres tablas con las que trabajamos. Los nombre de las tablas son:
 - Tabla_respuesta_gateway.csv
 - Tabla_provincias.csv
 - Tabla_comercio.csv
- **Directorio de ejecutables (bin por defecto):** contiene los scripts que forman parte de la inicialización y del proceso, estos son:
 - start.sh
 - stop.sh
 - inicializador.sh
 - proceso.sh
- **Directorio de novedades (nov por defecto)**
- **Directorio de aceptados (ok por defecto)**
- **Directorio de rechazados (nok por defecto)**
- **Directorio de procesados (proc por defecto)**
- **Directorio de salida (sal por defecto)**
- **BACKUPS:** este directorio guarda una copia de la forma original que tenía el directorio GRUPO03 antes de la instalación. Principalmente ayuda a la reparación de la instalación y si él hay archivos que no pueden ser recuperados.

2) Hipótesis y aclaraciones:

- La hora de cierre se define únicamente en el instalador. Una vez definidos los nombres de los directorios se le va a preguntar al usuario si quiere que la hora de cierre sea la hora por defecto o si va a querer que se cambie. Si el usuario elige esta última opción entonces el usuario tiene que elegir una hora en formato HHMMSS.
- La reparación la realiza el instalador únicamente. El instalador podrá recuperar cualquiera de los directorios que fueron creados en la instalación(**excepto BACKUPS**), cualquiera de los scripts que están en el directorio scripts y cualquiera de las tablas. Los archivos de log no se recuperan, se creará uno nuevo y se perderá toda la información previa. Si el archivo de configuración no se puede recuperar entonces el sistema se reinstalara, esto significa que todos los directorios creados antes serán destruidos se recupera la copia en BACKUPS y se lanza la instalación de nuevo.
- Todas las tablas están en formato .csv. Los archivos que el proceso evalúa y procesa tienen que tener extensión .txt, sino serán rechazados.
- Si el directorio de procesados no está porque se rompió la instalación y lanza el script inicializador o proceso entonces el archivo de log de ambos se crea en el directorio de ejecutables. Cuando se repare la instalación y se corran el proceso y el inicializador sus archivos de log correspondiente serán llevados al directorio de procesados.
- El proceso no puede lanzarse si el ambiente no es levantado. El inicializador cuando exporta las variables de ambiente crea una nueva que se llama HAYAMBIENTE y la setea en true. El script start es el que lanza el proceso, para verificar que el ambiente está levantado se fija si HAYAMBIENTE está seteado, si es así entonces start lanzará al proceso si no hay otro corriendo. Esto se realiza en las líneas 478 del inicializador y 284 del start.
- Si levanto el ambiente en una terminal y luego abro otra entonces podré levantar el ambiente en esa nueva terminal. Pero en esta nueva terminal no se me dejara correr el proceso si en la primera ya lo tengo corriendo.
- La verificación de que hay otro proceso corriendo la hace el script start, lo hace en la línea 290. Lo que hace es usar el comando **ps con la opción -C** pasándole como argumento el nombre del proceso. Aca se listan los procesos corriendo con ese nombre, si los hay. Luego con un **grep** se agarran las líneas de esa salida que terminen con la palabra "**proceso.sh**" y luego con el comando cut se agarra el primer valor de la salida, que es el process id. Si todos estos pasos dieron un resultado nulo, no hay proceso corriendo y lanza uno nuevo. si no es el caso le avisa al usuario que hay un proceso corriendo y se cierra el start.

- El proceso solo puede ser lanzado con el script start, este script también es llamado por el inicializador para lanzar al proceso si el usuario lo desea. Cualquier intento de lanzar al proceso manualmente por terminal resulta en la cancelación del mismo.
- Cuando se evalúan los registros de los archivos aceptados hay tres motivos por los cuales se los puede rechazar:
 - **Faltan campos:** Alguno de los campos de estos registros no esta, o tal vez no está en la posición que le corresponde.
 - **Formato inválido:** Alguno de los campos puede tener un formato que no le corresponde, por ejemplo un campo numérico tiene caracteres.
 - **Hay más de 15 registros.**
- Cuando se crea el registro de salida hay un campo llamado **"isO15_cResponseCodeShortDescription"**, este campo es la descripción corta de la tabla de respuestas de gateway para el código de respuesta que tiene el registro. En caso de que el código de respuesta del registro no esté en la tabla se le asignará el valor **"DESCONOCIDO"** al campo del registro de salida.
- Start escribe en el log del inicializador y stop en el del proceso.
- Si el directorio de novedades está vacío entonces el proceso duerme un minuto hasta el próximo ciclo, donde vuelve a hacer esta verificación. Solo cuando algún archivo llega al directorio de novedades y esté completa el minuto de suspensión el proceso empieza a procesar el archivo, y lo manda de novedades a aceptados y luego a procesados, o a rechazados, según corresponda. O sea que este proceso no deja archivos pendientes, los procesa todos en el mismo ciclo en las distintas etapas que les corresponden y crea la salida en el directorio de salida o en el de rechazados según corresponda.
- No se deben crear directorios nuevos a los que crea el sistema cuando se instala ya que pueden perderse de ser necesaria una reinstalación.

3) Descripción de problemas:

- El principal problema fue el poco conocimiento que se tenía de shell scripting, como el manejo de algunas estructuras como los vectores o usar el for para recorrer un archivo, entre otros, así como el funcionamiento de algunos de los comandos que se utilizaron. En general esto se resolvió yendo a buscar información sobre el problema o información sobre el comando en alguna página de internet o al campus de la materia.
- También en cómo hacer para evitar que dos procesos corran al mismo tiempo, la solución está explicada en la sección anterior.