
	PROCESO MEJORA E INNOVACION CARGUE DE INFORMACIÓN AUTOMATICA A STORAGE ACCOUNT EN AZURE	F3.P1.MI	09/06/2020
		Versión 4	Página 1 de 8

- 1. OBJETIVO:** Definir el cargue automático de nuevas solicitudes y reportes de vulneración de derechos a un Blob Storage para que estas sean leídas y procesadas por un cuadernillo de DataBricks en el cual se hace el Triage/Clasificación de acuerdo con el modelo de Machine Learning definido
- 2. ALCANCE:** Inicia con la programación del cargue de la información y culmina con la actualización de una tabla SQL que es consumida por un tablero de Power BI en el que los equipos de Defensoría de Familia pueden consultar el Triage/Clasificación de cada una de las solicitudes y reportes de vulneración de derechos.

3. DESCRIPCIÓN DE ACTIVIDADES:

No	Nombre de la Actividad	Descripción de la actividad	Responsable (Cargo o nivel que realiza la actividad)	Registro (evidencia de la actividad, si aplica)
		Inicio		
1	Confirmación de que nuevos registros han sido cargados	El área definida para compartir los nuevos registros de solicitudes y reportes de vulneración de derechos con el Grupo de Analítica Institucional confirma que el archivo del último día ha sido compartido en el lugar definido y con el nombre en formato "ddmmaaaa.txt", que corresponde al día, mes y año de los registros correspondientes (Por ejemplo, para los registros del 9 de agosto de 2021 el archivo debería llamarse "09082021.txt")	Dirección de Servicios y Atención ó Dirección de Información y Tecnología	
2	Programar tarea para archivo de fecha	<p>Programar la tarea en programador de tareas de Microsoft de creación de un archivo con la fecha actual, ya que los nombres de los nuevos archivos van a tener el nombre de fecha del día actual y por ello se debe crear un script que relacione correctamente el nombre del nuevo archivo a cargar, por eso creamos el siguiente archivo que se sobrescribe con la siguiente línea de comandos:</p> <p>DATE /T > "path.txt"</p> <p>Donde en "path" se define en donde se quiere que se aloje el archivo con la fecha del día actual que será posteriormente procesado.</p> <p>La tarea programada se crea de la siguiente manera:</p> <p>Se asigna un nombre y se configura de la siguiente manera:</p>		

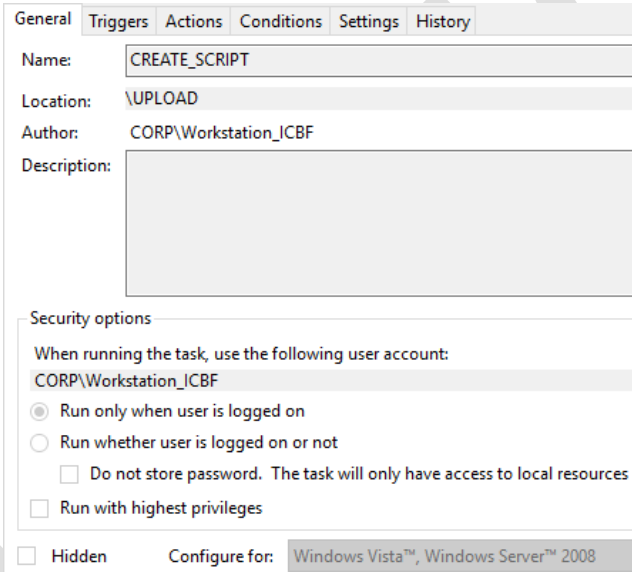
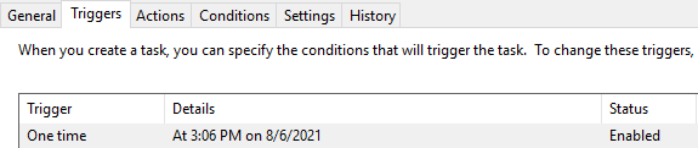
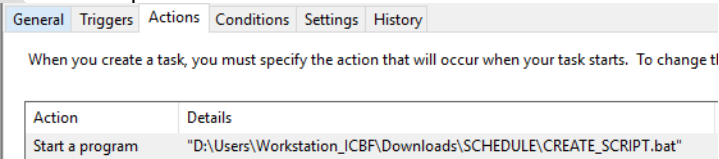
¡Antes de imprimir este documento... piense en el medio ambiente!

	<p style="text-align: center;">PROCESO MEJORA E INNOVACION</p> <p style="text-align: center;">CARGUE DE INFORMACIÓN AUTOMÁTICA A STORAGE ACCOUNT EN AZURE</p>	F3.P1.MI	09/06/2020
		Versión 4	Página 2 de 8

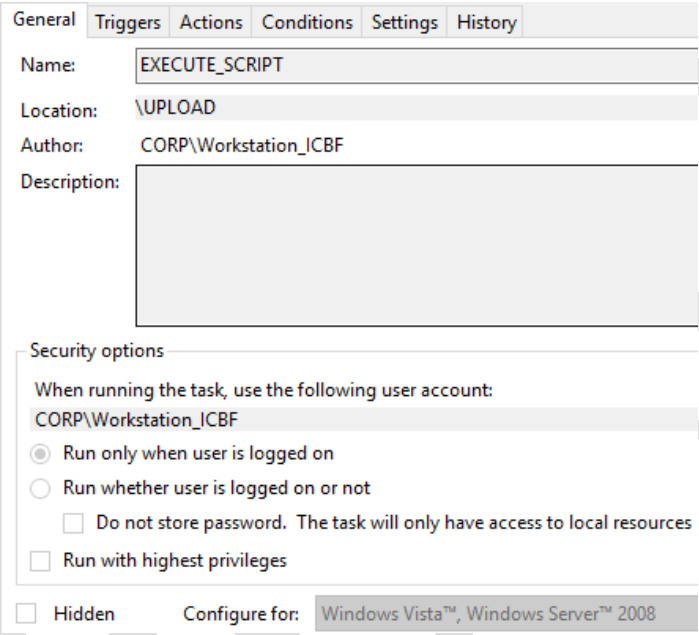
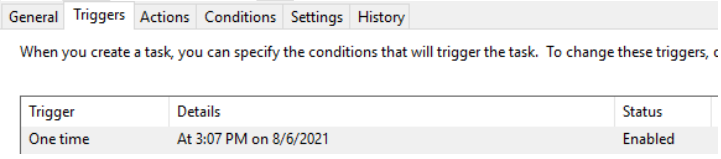
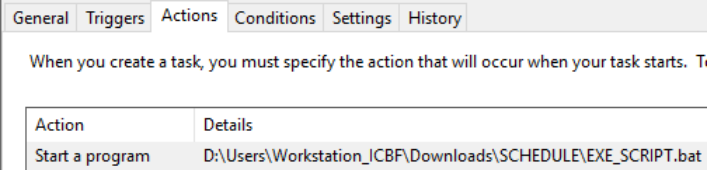
		<div><div><div>General</div><div>Triggers</div><div>Actions</div><div>Conditions</div><div>Settings</div><div>History</div></div><div><div>Name:</div><div>CREATE DATE</div></div><div><div>Location:</div><div>\\</div></div><div><div>Author:</div><div>CORP\Workstation_ICBF</div></div><div><div>Description:</div><div></div></div><div><div>Security options</div><div>When running the task, use the following user account: CORP\Workstation_ICBF</div><div><div><input type="radio"/> Run only when user is logged on</div><div><input checked="" type="radio"/> Run whether user is logged on or not</div><div><input type="checkbox"/> Do not store password. The task will only have access to local resources</div><div><input checked="" type="checkbox"/> Run with highest privileges</div></div><div><div><input type="checkbox"/> Hidden</div><div>Configure for: Windows Vista™, Windows Server™ 2008</div></div></div></div> <div>Se le asigna un disparador:</div> <div><div><div>General</div><div>Triggers</div><div>Actions</div><div>Conditions</div><div>Settings</div><div>History</div></div><div>When you create a task, you can specify the conditions that will trigger the task. To change these triggers, click the task.</div><div><table><tr><th>Trigger</th><th>Details</th><th>Status</th></tr><tr><td>One time</td><td>At 3:05 PM on 8/6/2021</td><td>Enabled</td></tr></table></div></div> <div>Y en la acción la ejecución del archivo en formato .bat con el comando expuesto arriba.</div> <div><div><div>General</div><div>Triggers</div><div>Actions</div><div>Conditions</div><div>Settings</div><div>History</div></div><div>When you create a task, you must specify the action that will occur when your task starts.</div><div><table><tr><th>Action</th><th>Details</th></tr><tr><td>Start a program</td><td>D:\Users\Workstation_ICBF\Downloads\SCHEDULE\GEN_DATE.bat</td></tr></table></div></div>	Trigger	Details	Status	One time	At 3:05 PM on 8/6/2021	Enabled	Action	Details	Start a program	D:\Users\Workstation_ICBF\Downloads\SCHEDULE\GEN_DATE.bat		
Trigger	Details	Status												
One time	At 3:05 PM on 8/6/2021	Enabled												
Action	Details													
Start a program	D:\Users\Workstation_ICBF\Downloads\SCHEDULE\GEN_DATE.bat													
3	Programar tarea de lectura de archivo de fecha y creación de script.	<p>Programar la tarea en programador de tareas de Microsoft para el procesamiento del documento creado con fecha del día actual y así crear el script de powershell con el cual mediante la herramienta nativa de Windows AZcopy cargamos el archivo del día a la cuenta de almacenamiento, para ello usamos el archivo creado en el paso anterior del cual extraemos los datos del número del día, mes y año, concatenamos con la ruta donde se encuentra el archivo a cargar y con la URL de destino (cuenta de almacenamiento) claramente siguiendo la sintaxis de la herramienta de AZCopy, esto con la siguiente línea de comandos:</p> <pre>@echo off setlocal enabledelayedexpansion for /F "tokens=*" %%a in (<path.txt >) do (SET linea=%%a SET mes=!linea:~4,-9!</pre>												

¡Antes de imprimir este documento... piense en el medio ambiente!






Cualquier copia impresa de este documento se considera como COPIA NO CONTROLADA.

		<pre>SET dia=!linea:~7,-6! SET agno=!linea:~10,-1! echo(<path>azcopy<path>!dia!!mes!!agno!.txt" "https://dlsicbfdenuncias.blob.core.windows.net/newdata?sp=<SAS>") > "<path>\SCRIPT.ps1"</pre> <p>Reemplazando <path> por cada una de las rutas de los correspondientes archivos. Y se programa la ejecución.</p> <p><SAS> por sus siglas en inglés (shared access signature) es una cadena de caracteres que se genera desde la cuenta de almacenamiento en el portal de Azure para autorizar la escritura de nuevos archivos y de esta manera proteger la seguridad de la cuenta de almacenamiento.</p> <p>La tarea programada se crea de la siguiente manera:</p> <p>Se asigna un nombre y se configura de la siguiente manera:</p>  <p>Se le asigna un disparador:</p>  <p>Y en la acción la ejecución del archivo en formato .bat con el comando expuesto arriba.</p> 		
3	Programar la	Programar la tarea en el programador de tareas de Microsoft para la		

¡Antes de imprimir este documento... piense en el medio ambiente!

<p>ejecución del script creado en la actividad 2</p>	<p>ejecución del script creado en la actividad 2, pues es necesaria una tarea que ejecute el script construido en el paso anterior y esto se logra creando un archivo que contenga la siguiente línea de comandos:</p> <p>Powershell -ExecutionPolicy Bypass -File "<path>\SCRIPT.ps1</p> <p>La tarea programada se crea de la siguiente manera:</p> <p>Se asigna un nombre y se configura de la siguiente manera</p>  <p><path> se refiere a la ruta donde quedó alojado el script hecho en la actividad 2.</p> <p>Se le asigna un disparador:</p>  <p>Y en la acción la ejecución del archivo en formato .bat con el comando expuesto arriba.</p> 		
<p>4 Programar el</p>	<p>Programar el disparador de la ejecución del cuadernillo de</p>		


disparador de la ejecución de cuadernillo	<p>DataBricks en Azure en el servicio de Azure Data Factory ya que este es el paso último que toma como insumo el archivo ya cargado. Esto se logra mediante el portal de Azure y con los siguientes pasos:</p> <p>Creamos un "linked service" que nos permite que nuestro servicio de Azure Data Factory tenga conexión a nuestro workspace de DataBricks donde se encuentra el cuadernillo a ejecutar para actualizar los registros con el archivo que ya hemos cargado previamente.</p> <div><p>Edit linked service (Azure Databricks)</p><p>Name * AzureDatabricks</p><p>Description <input type="text"/></p><p>Connect via integration runtime * ⓘ AutoResolveIntegrationRuntime</p><p>Account selection method * Enter manually</p><p>Databrick Workspace URL * ⓘ https://adb-6593758983612704.4.azure.databricks.net</p><p>Authentication type * Access Token</p><p>Access token * ⓘ *****</p><p>Select cluster</p><p>Save Cancel Test connection</p></div> <p>Y para ello solo necesitamos un token generado desde nuestro workspace de DataBricks que se obtiene dando clic en el icono como el nombre del Workspace.</p> <div><p>? dbw-icbf-denuncias</p><p>Signed in as luis.lizcano@icbf.gov.co</p><p>User Settings Admin Console Manage Account Log Out</p><p>Workspaces</p><p>databrickgai xotulio.ninox@icbf.gov.co</p><p>dbw-icbf-denuncias rodolfo.arteaga@icbf.gov.co</p><p>✓ dbw-icbf-denuncias luispa@bextsa.com</p><p>tdbricksicbfdb xotulio.mielese@icbf.gov.co</p></div> <p>Luego en User Settings y después en "Generate New Token".</p>		
---	--	--	--

	<div>User Settings</div> <div>Access Tokens Git Integration Notebook Settings Model Registry Settings Language Settings Preview</div> <div>Personal access tokens can be used for secure authentication to the Databricks API instead of passwords.</div> <div>Generate New Token</div> <p>Y el token que nos genera lo usaremos para crear nuestro “linked service”.</p> <p>Después de esto, en el área de trabajo de Azure Data Factory, creamos nuestro “pipeline” con una única actividad de Databricks que consiste en ejecutar el notebook “5_Prediccion_NuevosDatos”</p> <div><div>Notebook</div><div> Ejecucion_5_Prediccion_NuevosDatos</div><div>   </div></div> <p>A esta actividad se le asignan las siguientes opciones. En la pestaña “Azure DataBricks”.</p> <div><div>General Azure Databricks Settings User properties</div><div>Databricks linked service * <div>AzureDatabricks</div> Test connection Edit New</div></div> <p>Y en “Settings”.</p> <div><div>General Azure Databricks Settings User properties</div><div>Notebook path * <div>/Shared/5_Prediccion_NuevosDatos</div> Browse Open</div><div>Base parameters</div><div>Append libraries</div></div> <p>Ahora nada más nos resta crear un “Trigger” para que se ejecute el cuadernillo, teniendo en cuenta que su ejecución diaria se debe programar después un tiempo prudencial de al menos 5 minutos de la última actividad programada de cargue de archivos al Blob Storage.</p>		
--	---	--	--

	<div> <div>New trigger</div> <div> <div>Name *</div> <input type="text" value="trigger1"/> </div> <div> <div>Description</div> <input type="text" value="Execute trigger"/> </div> <div> <div>Type *</div> <div>Schedule</div> </div> <div> <div>Start date *</div> <input type="text" value="08/06/2021 8:09 PM"/> </div> <div> <div>Time zone *</div> <div>Coordinated Universal Time (UTC)</div> </div> <div> <div>Recurrence *</div> <div> <div>Every</div> <input type="text" value="1"/> <div>Day(s)</div> </div> </div> <div> <div>Advanced recurrence options</div> <div> <div>Execute at these times</div> <div> <div>Hours</div> <input type="text" value="11"/> </div> <div> <div>Minutes</div> <input type="text" value="59"/> </div> </div> </div> <div> <div>OK</div> <div>Cancel</div> </div> </div>		
	Fin		

P.C.: Punto de Control

4. **RESULTADO FINAL:** Se actualiza la tabla SQL que es consumida por Power BI para la actualización del tablero dirigido a los equipos de Defensoría de Familia con el Triage/Clasificación de las nuevas solicitudes y reportes que llegan al ICBF.
5. **DEFINICIONES:**
- AZcopy:** es una herramienta de línea de comandos que mueve datos dentro y fuera de Azure Storage.
 - Azure Data Factory:** es la plataforma que resuelve estos escenarios de datos. Se trata de un servicio de integración de datos y ETL basado en la nube que le permite crear flujos de trabajo orientados a datos a fin de coordinar el movimiento y la transformación de datos a escala.
 - Azure Databricks:** Azure Databricks es una plataforma de análisis de datos optimizada para la plataforma de servicios en la nube de Microsoft Azure. Azure Databricks ofrece tres entornos para desarrollar aplicaciones que consumen muchos datos: Databricks SQL, Databricks Data Science & Engineering y Databricks Machine Learning.
 - Linked Service:** Servicio ligado en español y hace referencia al servicio alojado en azure sobre el cual vamos a efectuar un procedimiento en este caso es el servicio de Databricks.
 - Workspace:** Espacio de trabajo en español y hace referencia a un espacio aislado del servicio de Databricks donde se alojan configuraciones de equipos de computo,

	PROCESO MEJORA E INNOVACION CARGUE DE INFORMACIÓN AUTOMÁTICA A STORAGE ACCOUNT EN AZURE	F3.P1.MI	09/06/2020
		Versión 4	Página 8 de 8

cuadernillos, repositorios, entre otros componentes necesarios para realizar procedimientos de lectura de información.

6. DOCUMENTOS DE REFERENCIA:

- **Sintaxis de AZcopy:**

<https://docs.microsoft.com/en-us/azure/storage/common/storage-ref-azcopy>

7. RELACIÓN DE FORMATOS (obligatorio): relacionar los formatos que son utilizados en la ejecución de las actividades del procedimiento a elaborar.

CODIGO	NOMBRE DEL FORMATO
No Aplica	No Aplica

8. CONTROL DE CAMBIOS (obligatorio): Registrar brevemente los cambios que se presenten en el cuerpo del documento según la versión que se modifica.

Fecha (Fecha de publicación de la versión del documento que se está reemplazando)	Versión (Relacionar la última versión y/o código del documento que se está Reemplazando)	Descripción del Cambio
09/08/2021	V1	No aplica.

Nota: En el caso de que el procedimiento sea nuevo (v1), se debe colocar en la tabla de control de cambios “**No Aplica**”.

¡Antes de imprimir este documento... piense en el medio ambiente!