



Facultad de Ingeniería
Departamento de Ingeniería en Sistemas
Bases de Datos II (IS-501)

Ing. Emilson Acosta

Integrantes:

- Abner Betancourt 20151021802
 - David Perdomo 20141031500
 - Leonardo Flores 20131006843
 - Nelson Díaz 20151001701
- Waleska Gutiérrez 20121003865

Ciudad Universitaria, Tegucigalpa MDC, Francisco Morazán

Abril, 2018

Índice:

1. Introducción

2. Flipboard

3. Base de Datos OLTP

4. Base de Datos OLAP

5. Reportes con Pentaho

6. Replicación de Alta Disponibilidad

1. Introducción

Desde una editorial de prensa que sirve a modo de resumen de lo que solemos perder de continuidad en una afanada semana, hasta aquella sensacionalista “prensa rosa” que tanto cotilleo causa en las portadas de los diferentes rotativos semanales. Los portales interactivos noticieros son una fuente de tráfico en Internet que genera numerosas reacciones en base a su contenido disponible... Las plataformas informativas han desplazado en un gran porcentaje a los medios convencionales debido a su facilidad de acceso y su nulo costo de adquisición. Puesto que sus ganancias son generadas por el tráfico promedio que establece las tasas de publicidad presentadas. Cabe de destacar la reciente concientización que surge no solo de la fuente informativa, sino también de la calidad del formato investigativo, periodístico y hasta su formato de presentación (prensa amarillista, sensacionalista, etcétera). Lo importante aquí, es el análisis desde el punto de vista web y en base a su diseño relacionado a las bases de datos que conforman el portal informativo Flipboard.

Se ha desarrollado un proyecto de DATAWAREHOUSE de una empresa ficticia. Un data warehouse es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso, para esto contamos con una base de datos OLTP que tiene suficiente información almacenada y que permite la generación de reportes mediante el uso de técnicas de Inteligencia de Negocios vistas en clase.

Hemos utilizado una base de datos OLTP de una pagina de una empresa de agregación de noticias y redes sociales con sede en Palo Alto, California, con oficinas en Nueva York, Vancouver y Beijing. Nuestra base de datos OLTP es de tipo de esquema estrella; un esquema en estrella es un modelo de datos que tiene una tabla de hechos que contiene los datos para el análisis, rodeada de las tablas de dimensiones. Este aspecto, de tabla de hechos (o central) más grande rodeada de radios o tablas más pequeñas es lo que asemeja a una estrella, dándole nombre a este tipo de construcción.

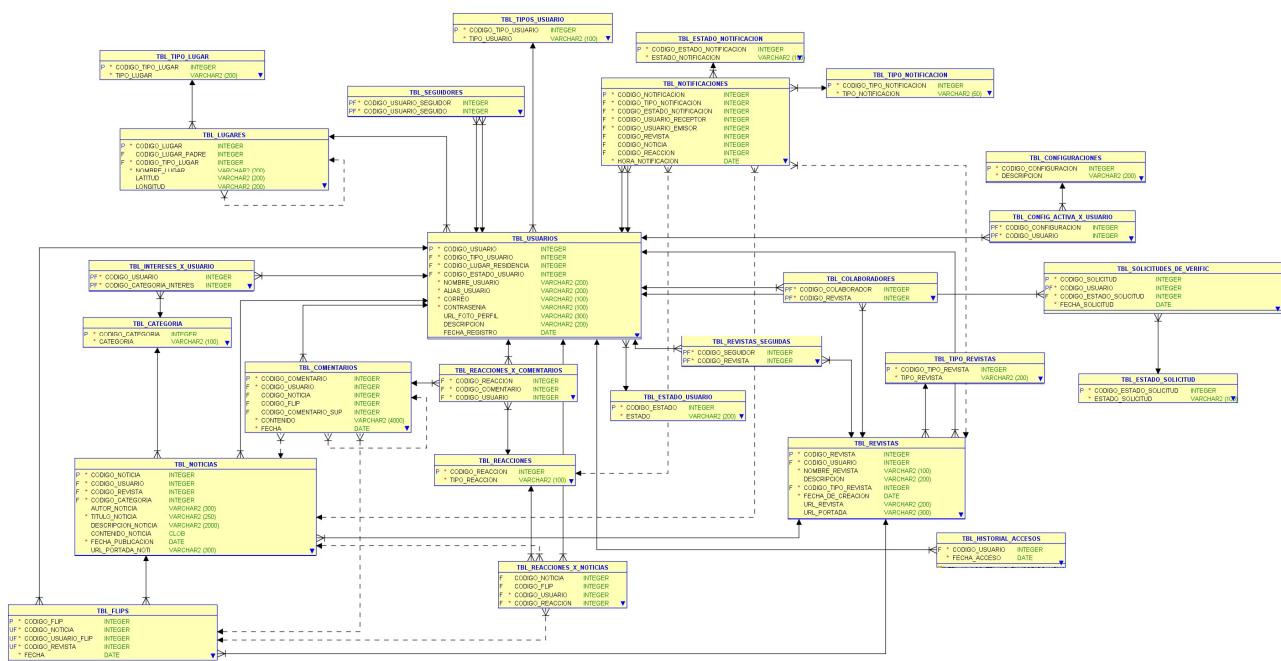
El proceso de trabajo del Proyecto comenzó al construir la base de datos y crear los numerosos registros dentro de la misma, todos los integrantes del grupo comenzaron a crear registros en la base de datos para que tuviera información relevante dentro de ella para poder utilizar la información.

2. Flipboard

Flipboard es una empresa de agregación de noticias y redes sociales con sede en Palo Alto, California, con oficinas en Nueva York, Vancouver y Beijing. Su software, también conocido como Flipboard, se lanzó por primera vez en julio de 2010. Se agrega el contenido de las redes sociales, noticias, sitios de intercambio de fotos y otros sitios web, lo presenta en formato de revista, y permite a los usuarios "flipear" a través de los artículos, imágenes y videos que se comparten. Los lectores también pueden guardar historias en revistas de Flipboard. A partir de marzo de 2016 la compañía afirma que ha habido 28 millones de revistas creadas por los usuarios en Flipboard. La compañía recaudó más de \$ 200 millones en fondos de inversionistas, incluyendo Kleiner Perkins Caufield & Byers, Index Ventures, Rizvi Traverse Management y Insight Venture Partners, y \$ 50 millones adicionales de JPMorgan Chase en julio de 2015. En marzo de 2014, Flipboard compró Zite, una aplicación de lectura de revista, de la cadena de televisión CNN. El filtrado de contenidos de Flipboard, el motor de temas y el sistema de recomendaciones se integraron a partir de esta adquisición. Zite fue cerrado el 7 de diciembre de 2015. En febrero de 2015, Flipboard se hizo disponible en la web. Hasta entonces, Flipboard era una aplicación para móviles, sólo disponible en tabletas y teléfonos móviles. El cliente web proporciona enlaces a páginas web en los navegadores de escritorio y carece de algunas características del software cliente. En febrero de 2017, Flipboard actualizó sus aplicaciones móviles para iOS y Android a 4.0, lo que trajo un rediseño completo de la aplicación, e implementó nuevas características como revistas inteligentes, que permiten a los usuarios agrupar diferentes cosas, como varias fuentes de noticias, personas y hashtags. Flipboard genera ganancias de diferentes anunciantes que crean sus revistas en base a lo que represente a su marca. Estos pueden crear noticias propias, que son añadidas a sus revistas.

3. Bases de Datos OLTP

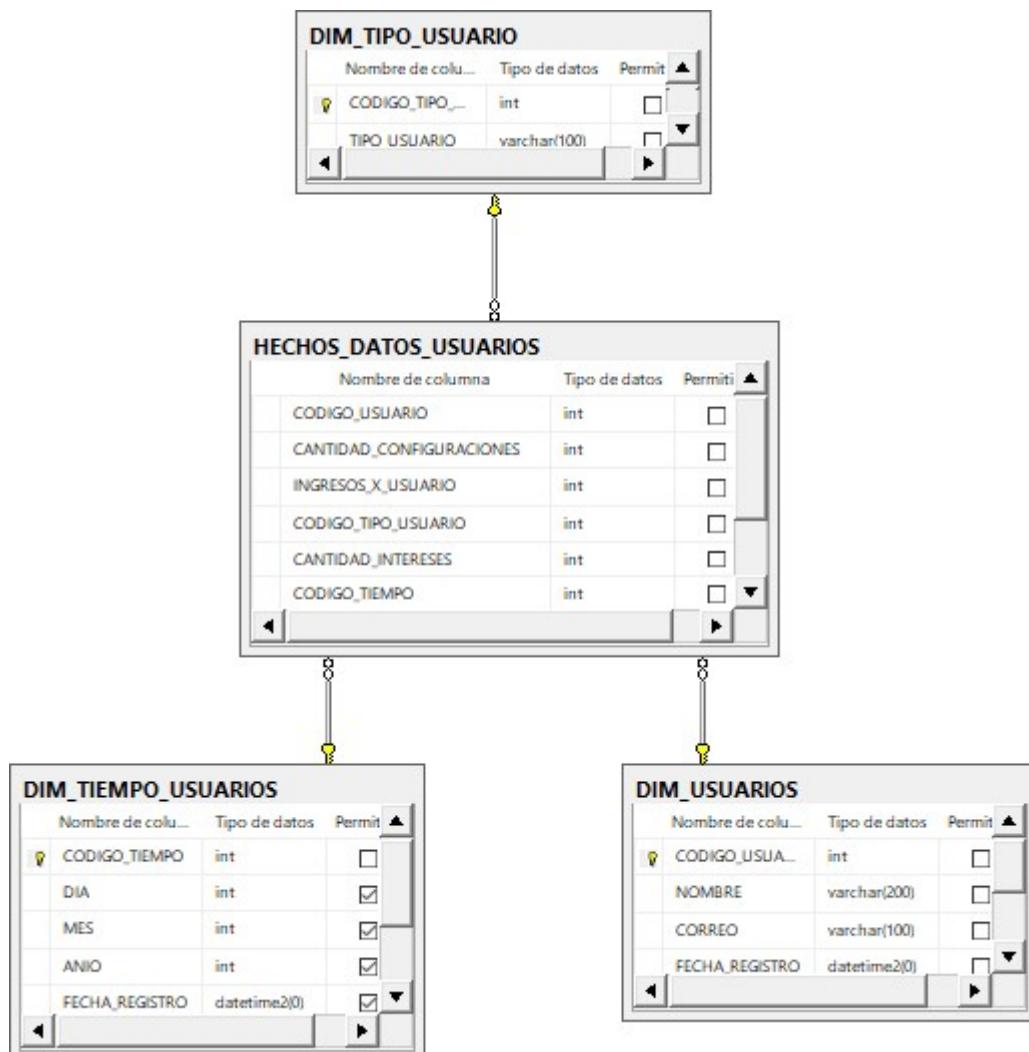
Las bases de datos OLTP están orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico (que debe ser validado con un commit, o invalidado con un rollback), y que puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales.



Nota La imagen se encuentra en mejor resolución en GitHub.com/ BD2-I2018/Bases de Datos/Oracle/Esquema.png

4. Bases de Datos OLAP

Los sistemas OLAP son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos... etc. Este sistema es típico de los datamarts.

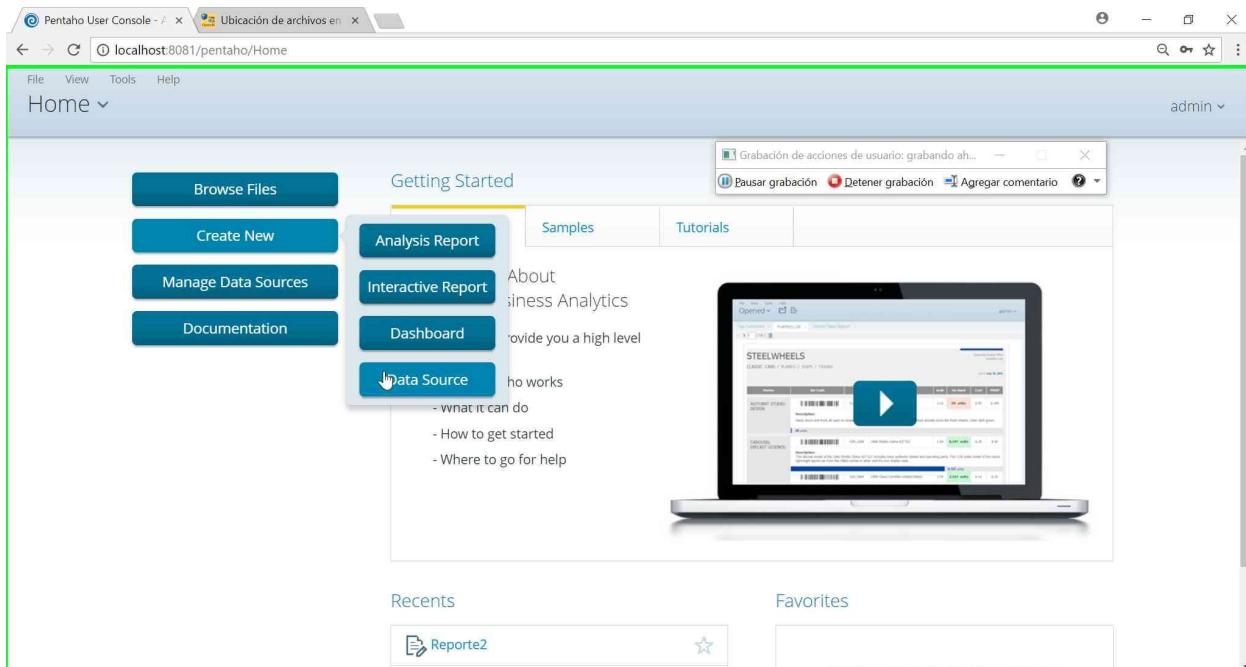


Nota La imagen se encuentra en mejor resolución en [GitHub.com/ BD2-I2018/Bases de Datos/SQL Server/OLAP/Esquema.png](https://github.com/BD2-I2018/Bases_de_Datos_SQL_Server_OLAP_Esquema.png)

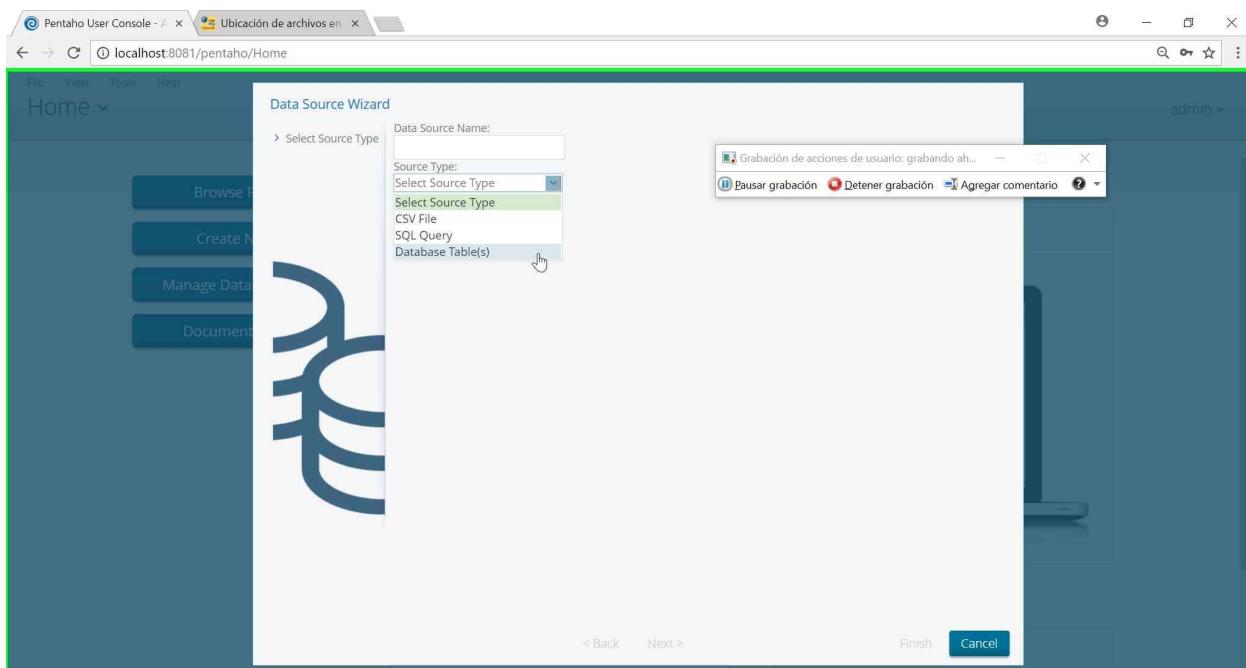
5. Reportes con PENTaho

Pentaho, que es una compañía de software business intelligence que ofrece productos open source que nos otorga integración de datos, servicios OLAP, reportes, tableros de información, data mining y capacidades ETL.

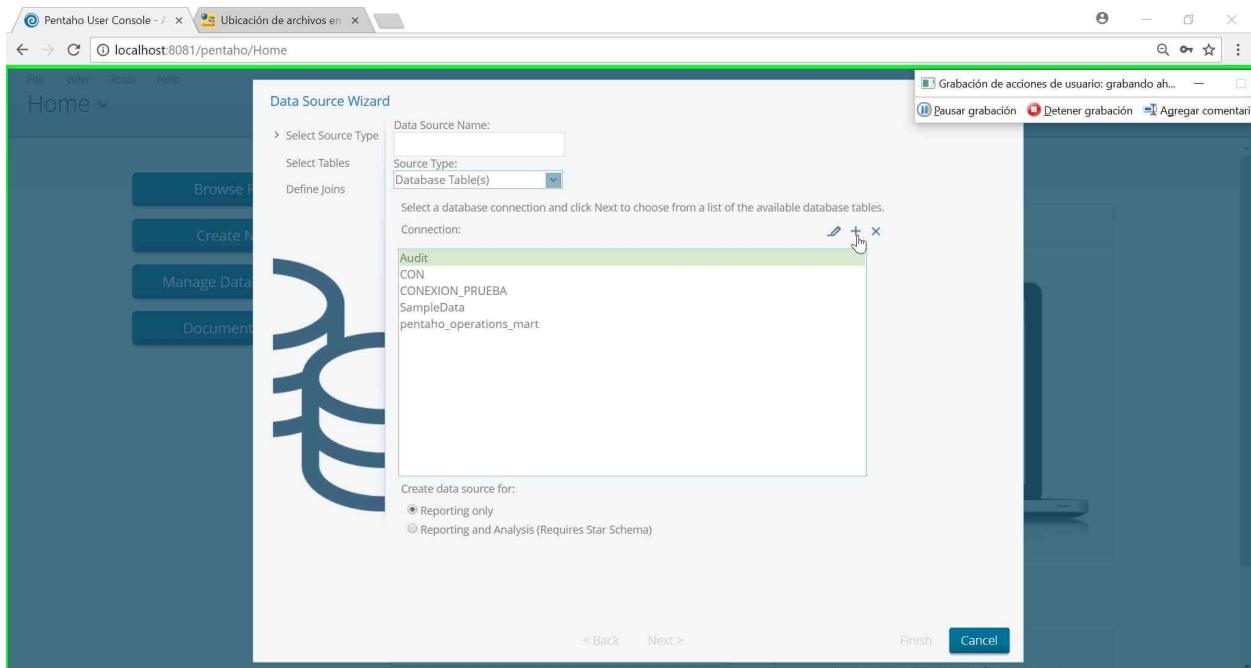
1. Creamos un nuevo data source:



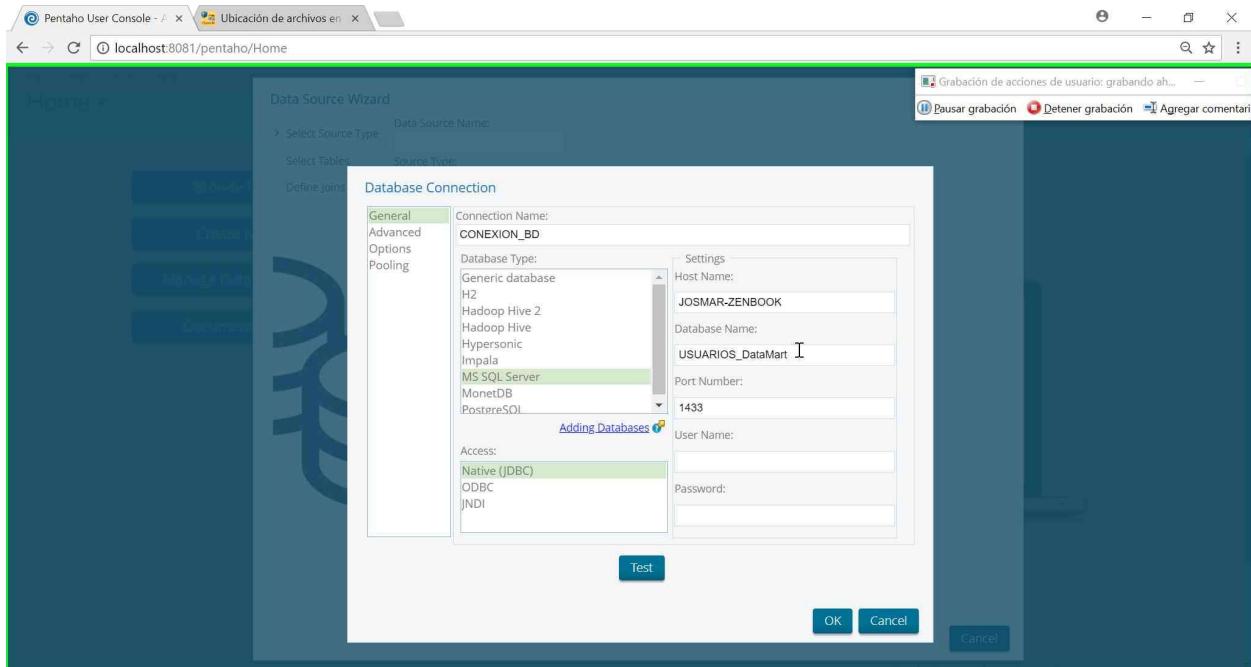
2. En la ventana emergente seleccionamos Database Table(s)



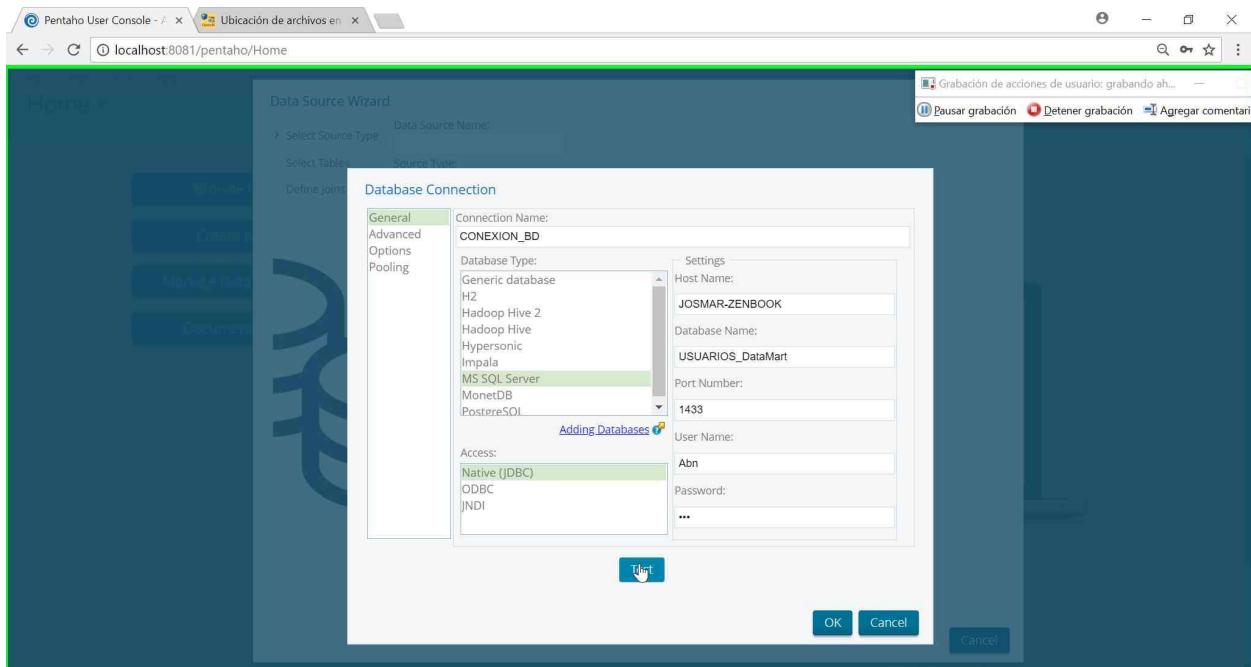
3. Antes de crear el data source agregamos una nueva conexión (+).



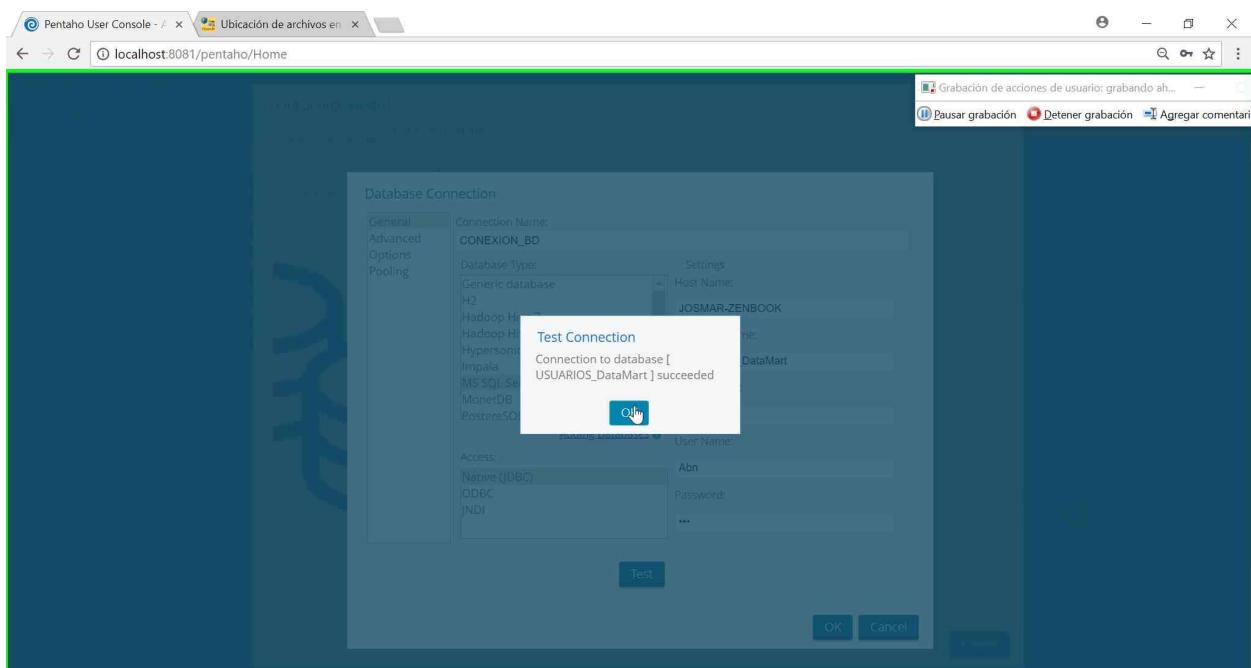
4. Seleccionamos que tipo de motor o gestor vamos a conectar, en nuestro caso 'MS SQL Server', agregamos el nombre de Host, nombre de Base de datos, configuramos el puerto (por defecto 1433) e ingresamos con un usuario valido del gestor.



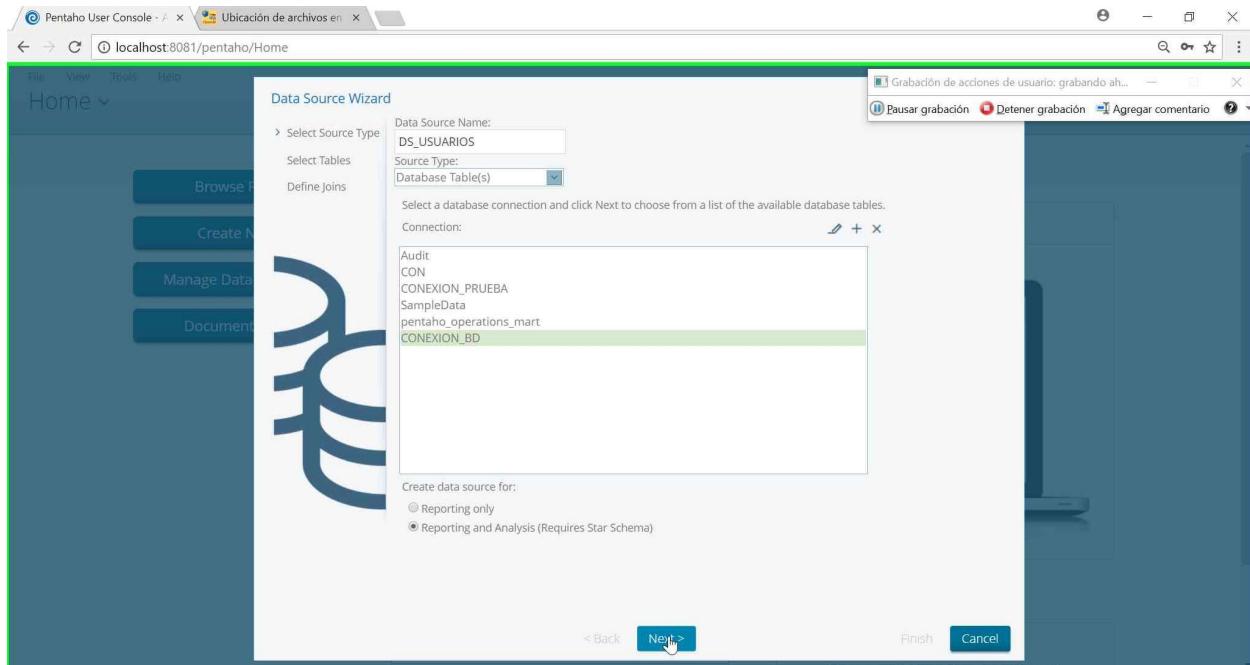
5. Ahora probaremos la conexión



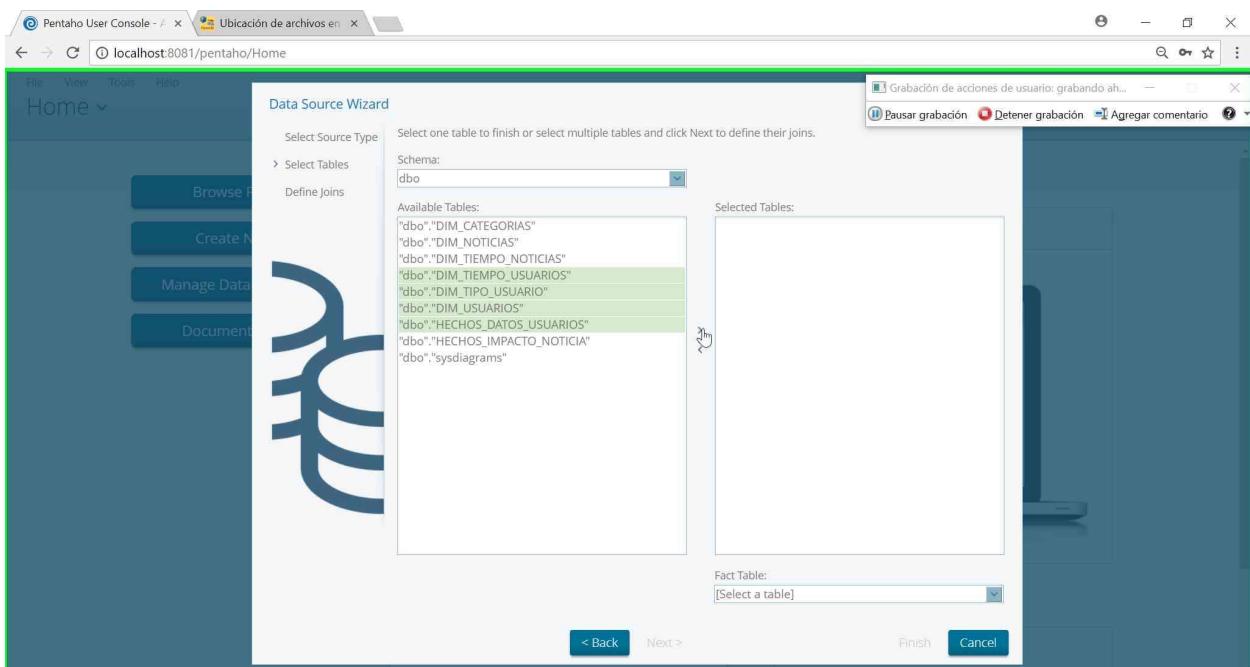
6. Como la conexión ha sido un éxito, procederemos a continuar



7. Le asignamos un nombre a nuestro Data Source y seleccionamos la conexión que acabamos de crear (Tambien especificamos en la parte de abajo que es para reporte y análisis)



8. Seleccionamos las tablas con las que trabajaremos y las agregaremos en el espacio del lado derecho.



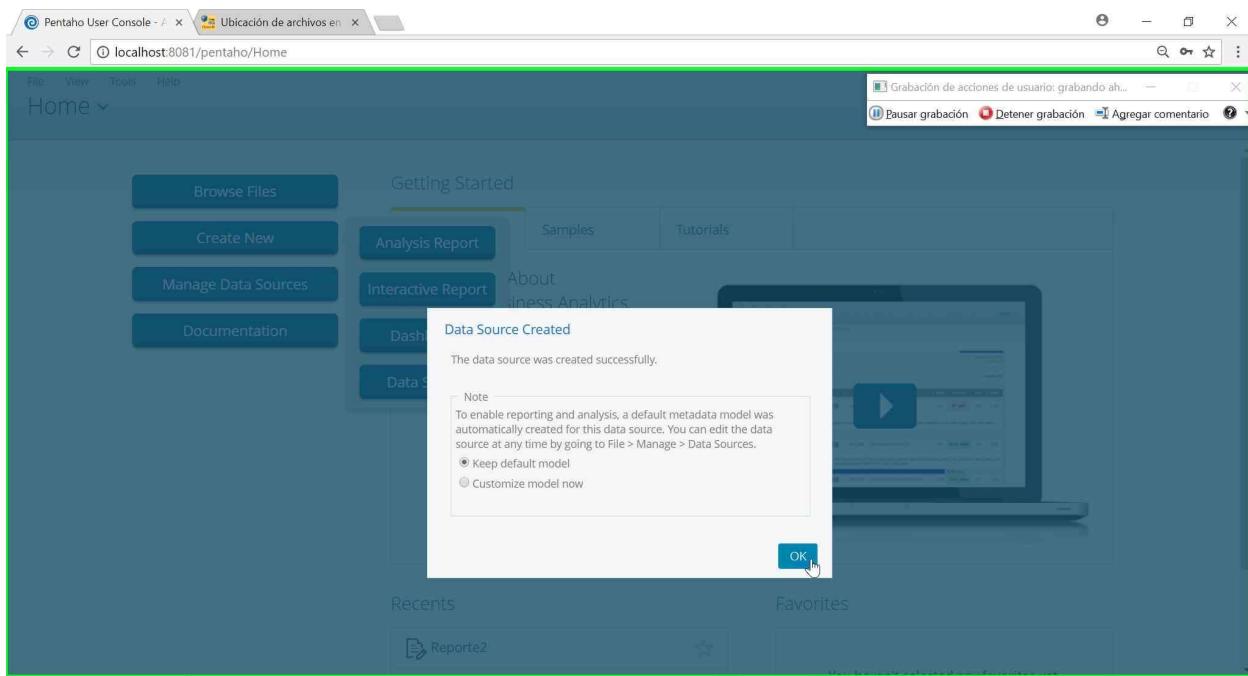
9. Una vez terminado eso, seleccionamos la tabla de hechos

The screenshot shows the 'Data Source Wizard' interface in the Pentaho User Console. The title bar says 'Ubicación de archivos en' and 'localhost:8081/pentaho/Home'. The main panel is titled 'Data Source Wizard' and has a sub-section 'Select Tables'. A dropdown 'Schema:' is set to 'dbo'. Below it is a list of 'Available Tables': 'dbo"."DIM_CATEGORIAS"', 'dbo"."DIM_NOTICIAS"', 'dbo"."DIM_TIEMPO_NOTICIAS"', 'dbo"."HECHOS_IMPACTO_NOTICIA"', 'dbo"."sysdiagrams"', and 'dbo"."HECHOS_DATOS_USUARIOS"'. To the right, under 'Selected Tables:', are 'dbo"."DIM_TIEMPO_USUARIOS"', 'dbo"."DIM_TIPO_USUARIO"', 'dbo"."DIM_USUARIOS"', and 'dbo"."HECHOS_DATOS_USUARIOS"'. At the bottom, there's a 'Fact Table:' section with a dropdown set to '[Select a table]' and a list of tables: 'dbo"."HECHOS_DATOS_USUARIOS"', 'dbo"."DIM_TIPO_USUARIO"', 'dbo"."DIM_USUARIOS"', and 'dbo"."HECHOS_DATOS_USUARIOS"'. Navigation buttons at the bottom include '< Back', 'Next >', and 'Finish'.

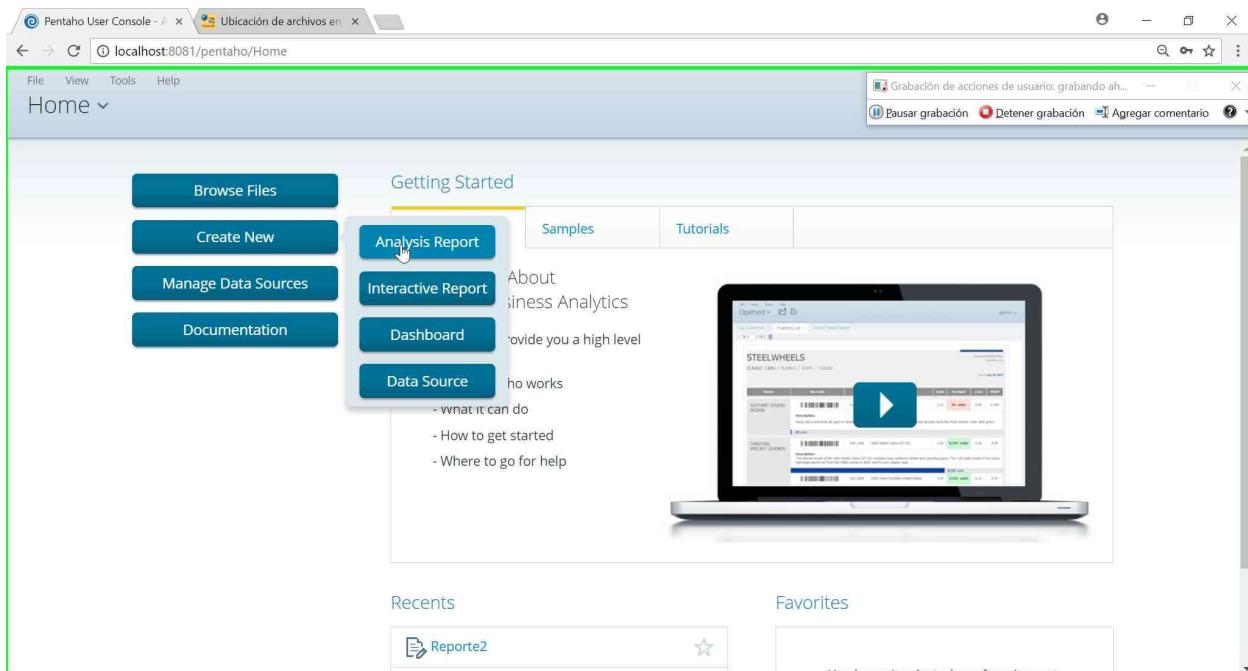
10. Agregamos los join de la tabla de hechos que estan relacionados con las demás tablas que agregamos, y finalizamos nuestro Data Source.

The screenshot shows the 'Data Source Wizard' interface in the Pentaho User Console. The title bar says 'Ubicación de archivos en' and 'localhost:8081/pentaho/Home'. The main panel is titled 'Data Source Wizard' and has a sub-section 'Define Joins'. It shows 'Left Table:' as 'dbo"."HECHOS_DATOS_USUARIOS"' and 'Right Table:' as 'dbo"."DIM_USUARIOS"'. Under 'Key Field:', 'CÓDIGO_USUARIO' is selected from the left and 'CÓDIGO_USUARIO' is selected from the right. A 'Create join' button is visible. Below, the 'Join(s):' section lists three joins: 'dbo"."HECHOS_DATOS_USUARIOS".CÓDIGO_TIEMPO - INNER JOIN - "dbo"."DIM_TIEMPO_USUARIO"', 'dbo"."HECHOS_DATOS_USUARIOS".CÓDIGO_TIPO_USUARIO - INNER JOIN - "dbo"."DIM_TIPO_USUARIO"', and 'dbo"."HECHOS_DATOS_USUARIOS".CÓDIGO_USUARIO - INNER JOIN - "dbo"."DIM_USUARIOS"'. Navigation buttons at the bottom include '< Back', 'Next >', 'Finish', and 'Cancel'.

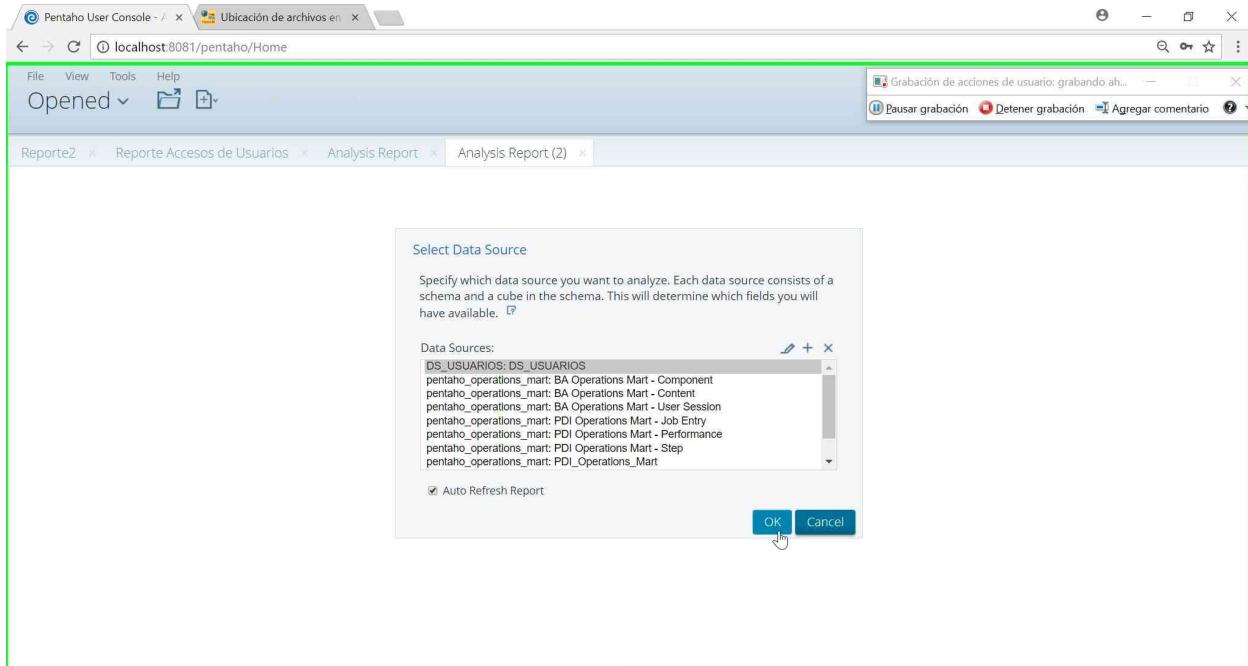
11. Dejamos la opción por defecto y continuamos con la creación de reportes.



12. Nos iremos a Create New -> Analysis Report



13. Seleccionamos el Data Source que creamos



14. Ahora simplemente arrastrando los campos que queramos en los espacios de 'Rows', 'Columns' y 'Measures' Creamos el reporte.

The screenshot shows the 'Analysis Report (2)' interface in the Pentaho User Console. On the left, there's a sidebar titled 'Available fields (17) for: DS_USUARIOS' with a list of fields including 'NOMBRE', 'CÓDIGO TIEMPO', 'CÓDIGO TIPO USUARIO', 'CÓDIGO USUARIO', 'INGRESOS X USUARIO', and others. The main area is a pivot table with the following structure:

CÓDIGO U	NOMBRE	MES							
		1	3	4	5	6	7	8	9
	INGRESOS X USUARIO								
1	Administrador	-	-	-	-	-	-	-	
2	Marco Polo	-	-	-	-	-	-	-	
3	Rosa Betancourt	-	-	-	-	-	-	-	
4	Abdel Hadi	-	-	-	-	-	-	-	
5	Luis Messen	-	-	-	-	-	-	-	
6	Ludivine Madje	-	-	-	-	-	-	-	
7	Marie Valeria	-	-	-	-	-	-	-	
8	Andy Weith	-	-	-	-	-	-	-	
9	Bea Ramos	5	-	-	-	-	-	-	
10	Allan Orellana	-	-	-	-	-	-	-	
11	Pedro Suntecan	6	-	-	-	-	-	-	
12	Clara Castrati	-	-	-	-	-	-	-	
13	Sara Donedu	-	-	-	-	-	-	-	
14	Abel Vera	-	-	-	-	-	-	-	
15	Abel Angeles	-	-	-	-	-	-	-	
16	Carla Patricia	-	-	-	-	-	-	-	
17	Charles Manson	-	-	-	-	-	-	-	
18	Mauricio Matamoros	-	-	-	-	-	-	-	
19	Enrique Lardizabal	-	-	-	-	-	-	-	
20	Patricia Fernandez	-	-	-	-	-	-	5	

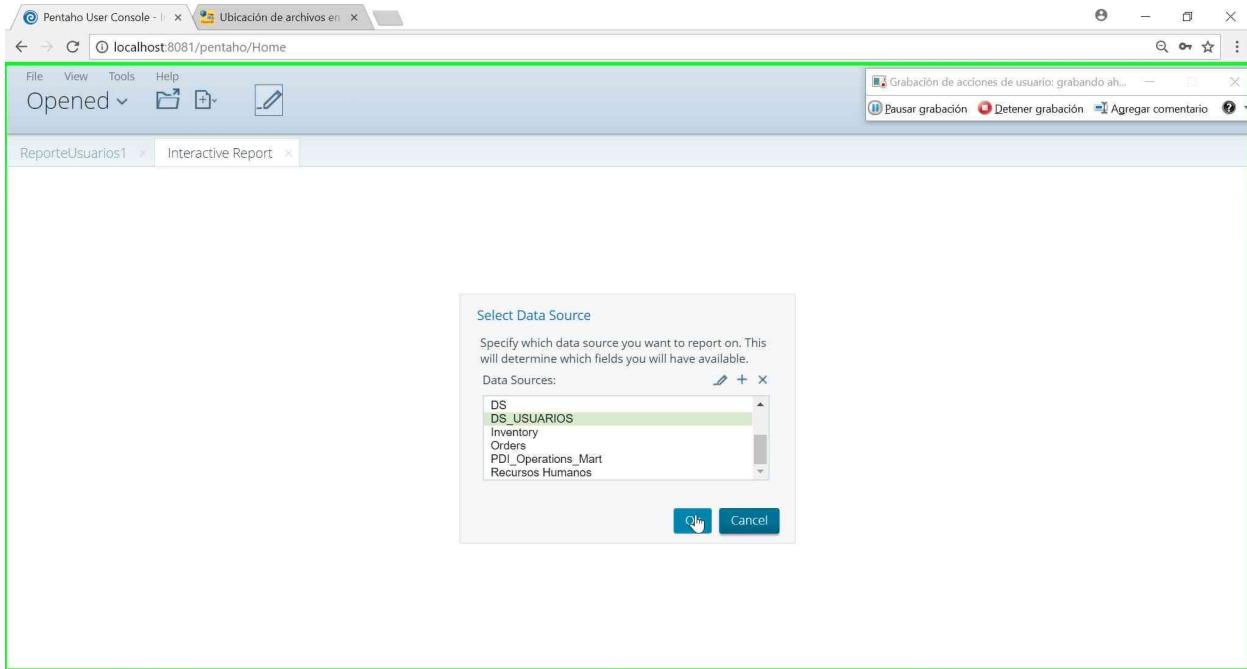
15. Una vez finalizado el reporte que queremos, procederemos a guardarlo.

The screenshot shows the Pentaho User Console interface. A 'Save' dialog box is open in the foreground, prompting the user to save a report named 'ReporteUsuarios' to the location '/home/admin'. The background displays the Analysis Report configuration screen, which includes a sidebar with available fields like 'CODIGO USUARIO', 'NOMBRE', 'ANIO', 'MES', and 'INGRESOS X USUARIO'. The main area shows a grid with columns for 'ANIO' and 'MES' (with values 2012, 5, 10, 1, 4, 8, 10) and rows for various users. The status bar at the top indicates 'Grabación de acciones de usuario: grabando ah...'.

16. Para crear un reporte interactivo, accederemos en la parte de menu en la pestaña File, New y por ultimo Interactive Report.

The screenshot shows the Pentaho User Console interface with the 'File' menu open. The 'Interactive Report' option is selected, indicating the creation of an interactive report. The main workspace shows the Analysis Report configuration screen, which includes a sidebar with available fields like 'CODIGO USUARIO', 'NOMBRE', 'ANIO', 'MES', and 'INGRESOS X USUARIO'. The main area shows a grid with columns for 'ANIO' and 'MES' (with values 2012, 5, 10, 1, 4, 8, 10) and rows for various users. The status bar at the top indicates 'Grabación de acciones de usuario: grabando ah...'.

17. Nuevamente seleccionamos nuestro Data Source

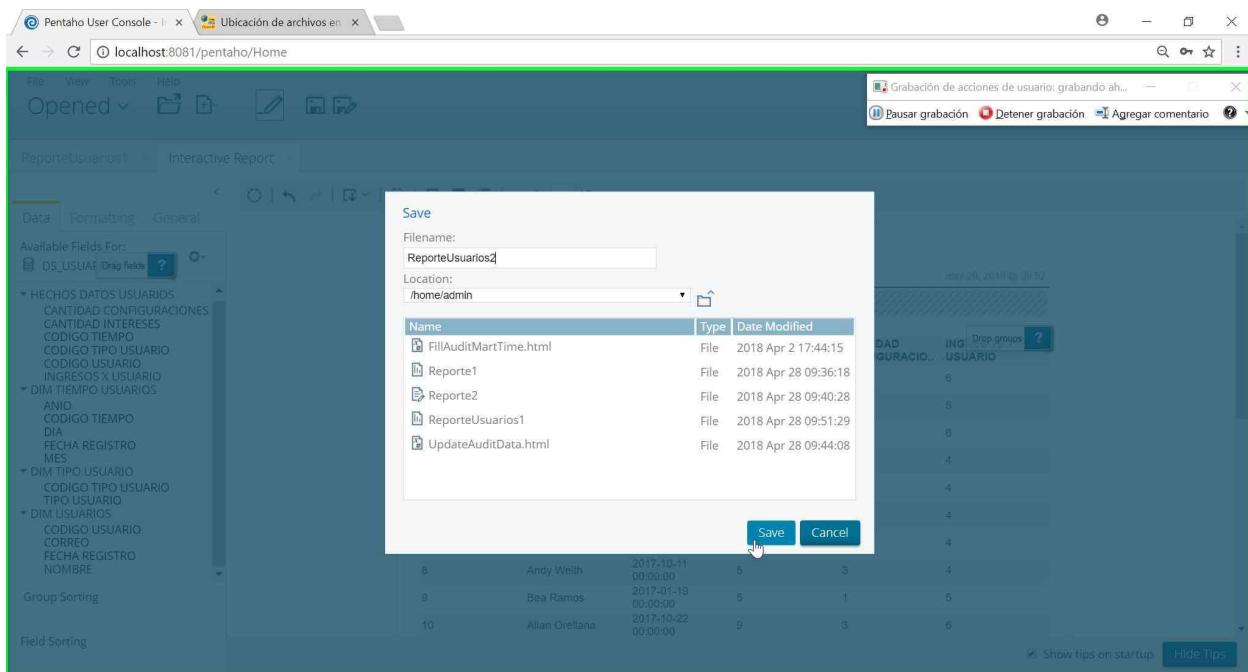


18. De la misma manera, simplemente arrastrando los campos deseados, creamos el reporte.

The screenshot shows the Pentaho User Console interface with a report titled 'Reporte DE USUARIOS'. On the left, there is a sidebar titled 'Available Fields For:' with a tree view of data sources. The 'DS DS_USUARIOS' node is expanded, showing fields like 'HECHOS DATOS USUARIOS', 'DIM TIEMPO USUARIOS', 'DIM TIPO USUARIO', 'DIM USUARIOS', and 'Group Sorting'. The main area displays a table with the following data:

CODIGO USUARIO	NOMBRE	FECHA REGISTRO	CANTIDAD INTERESES	CANTIDAD CONFIGURACIONES	INGRUSOS X USUARIO
1	Administrador	2017-10-19 00:00:00	4	1	6
2	Marco Polo	2017-10-19 00:00:00	0	3	5
3	Rosa Betancourt	2017-10-19 00:00:00	6	4	6
4	Abdel Hadi	2017-10-19 00:00:00	4	3	4
5	Luis Messen	2017-11-19 00:00:00	6	3	4
6	Ludivine Madej	2018-10-19 00:00:00	6	2	4
7	Marie Valeria	2017-12-10 00:00:00	7	2	4
8	Andy Weith	2017-10-11 00:00:00	5	3	4
9	Bea Ramos	2017-01-19 00:00:00	5	1	5
10	Alian Orellana	2017-10-22 00:00:00	9	3	6

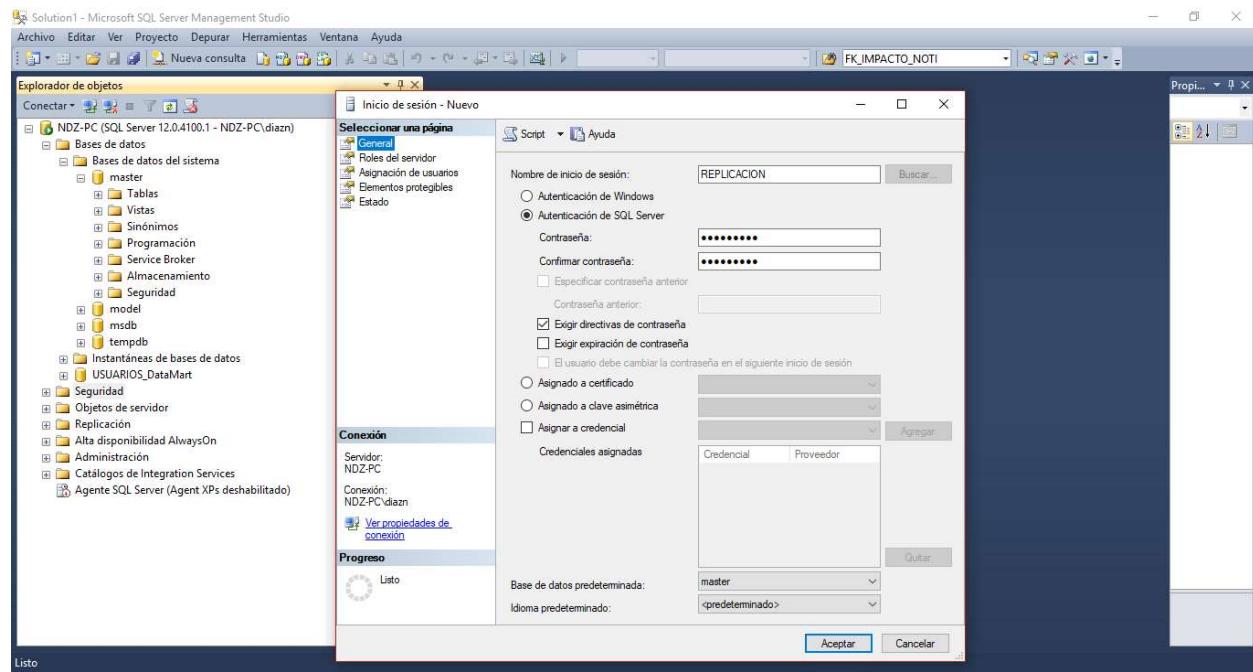
19. Finalmente guardamos el reporte cuando consideremos que esta listo.



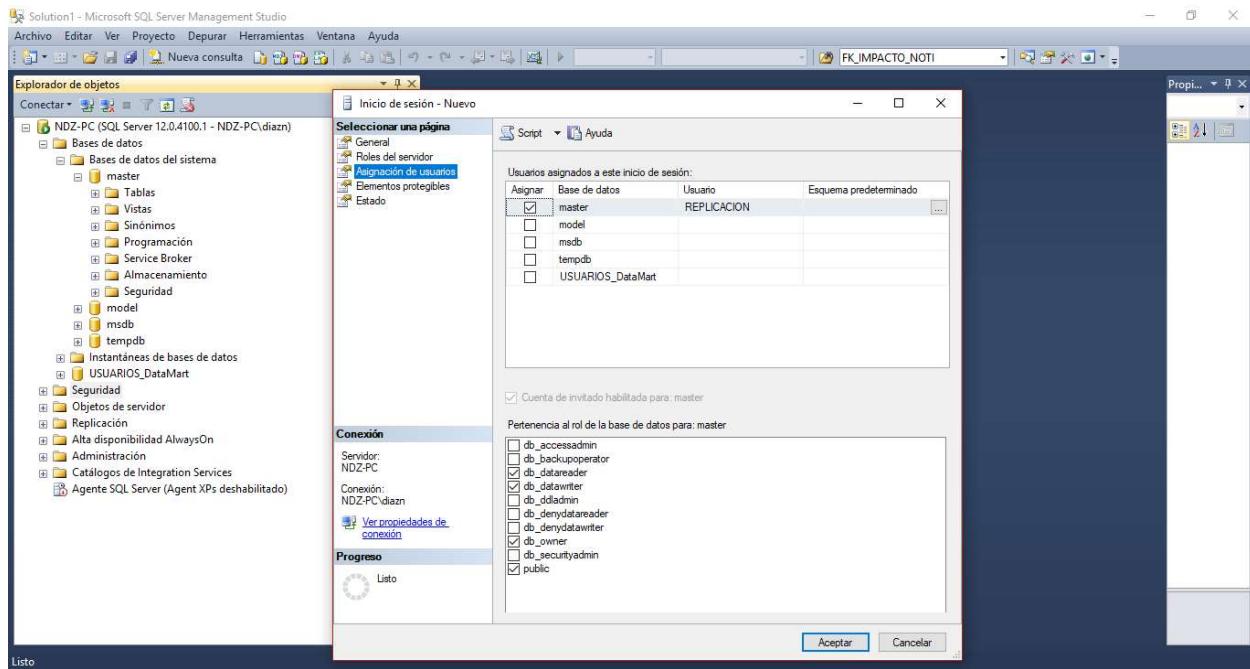
6. Replicación en Alta Disponibilidad

La replicación utiliza un modelo de publicación y suscripción, lo que permite que un servidor principal, conocido como publicador, distribuya datos a uno o varios servidores secundarios o suscriptores. Gracias a la replicación se puede obtener una disponibilidad y escalabilidad en tiempo real entre tales servidores. Esta solución admite el filtrado para obtener un subconjunto de datos de los Suscriptores, así como las actualizaciones con particiones. Los suscriptores están en línea y disponibles para la generación de informes y otras funciones, sin recuperación de consultas. SQL Server ofrece tres tipos de replicación: replicación de instantáneas, replicación transaccional y replicación de mezcla. La replicación transaccional proporciona la latencia más baja y es la que más se suele utilizar por su alta disponibilidad. Para obtener más información, vea Mejorar la escalabilidad y la disponibilidad.

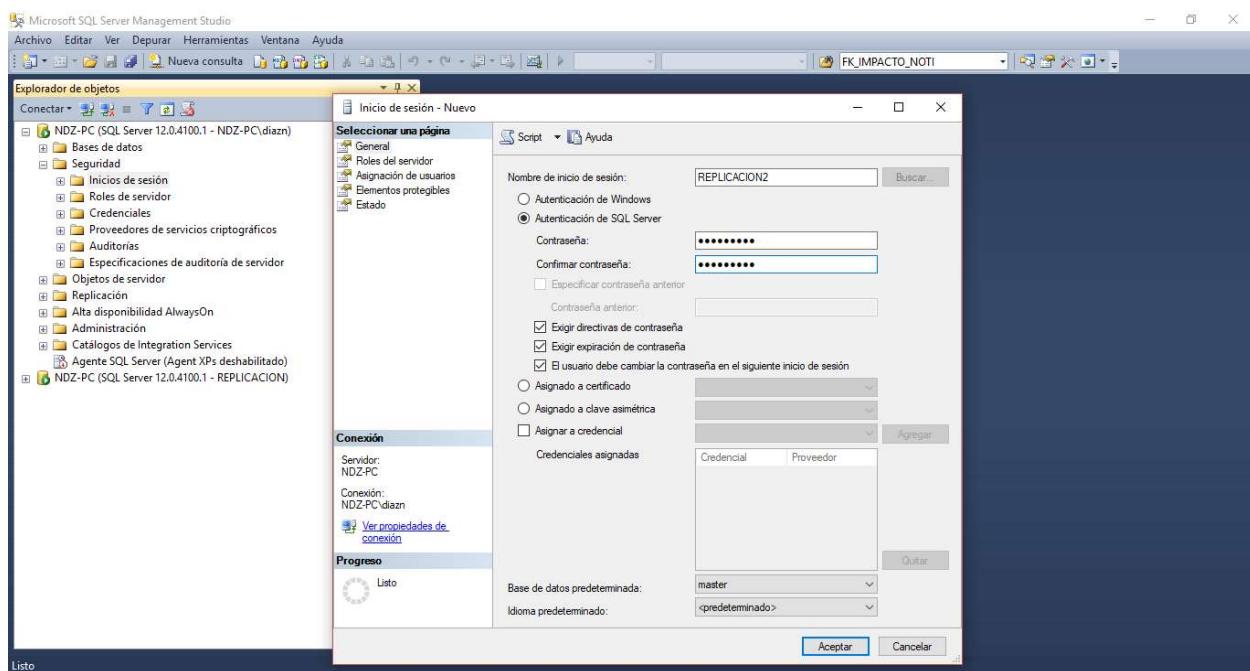
1. Se crea un usuario para la replicacion llamado "REPLICACION" con clave acceso "sqlserver".



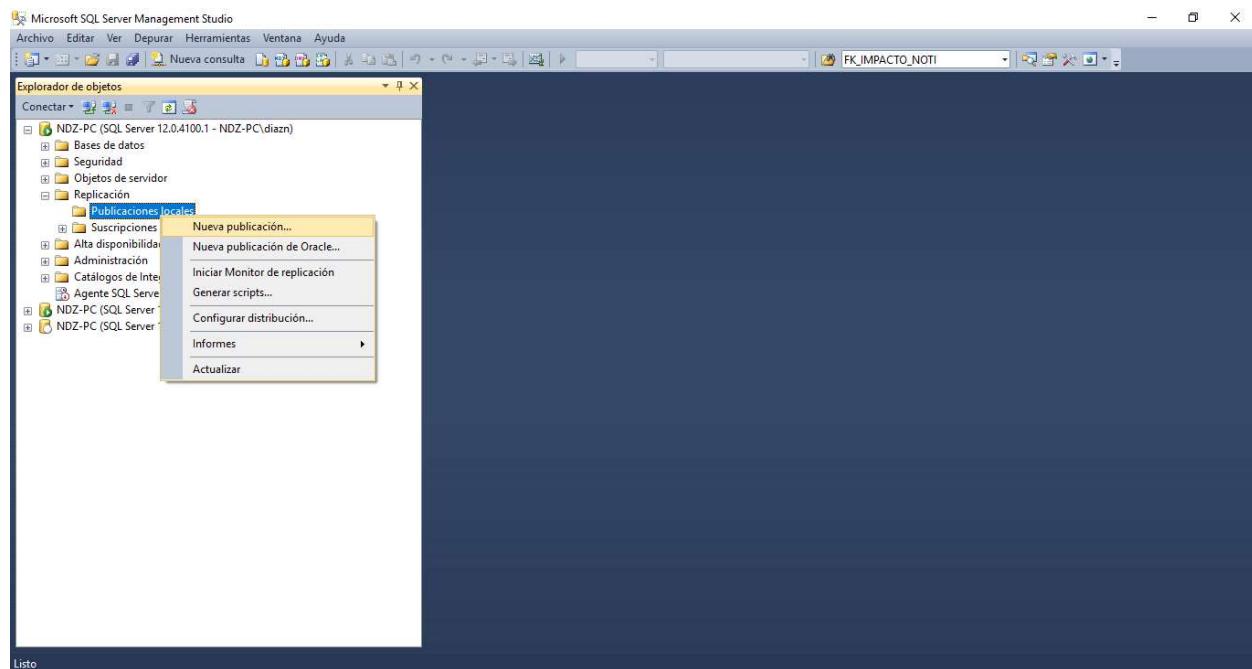
2. Se le asignan los privilegios (al usuario antes creado) de lectura, escritura y propiedad sobre la base de datos a replicar.



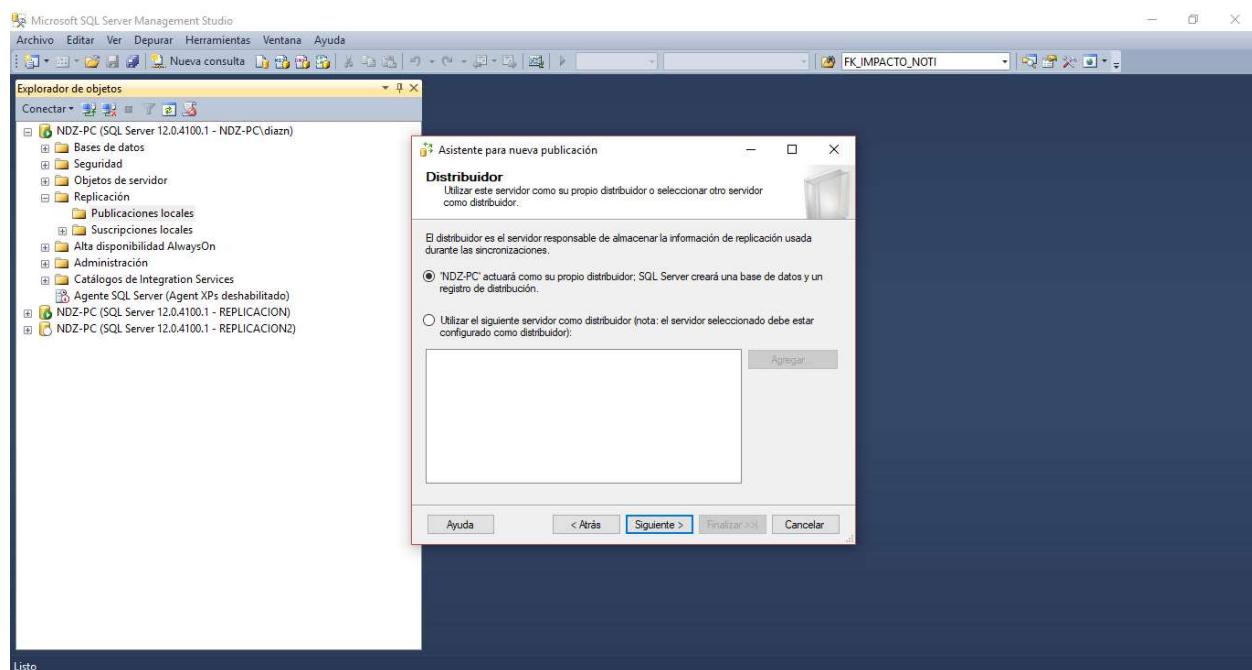
3. Repetimos los pasos 1 y 2 para otro usuario llamado "REPLICACION2" con clave de acceso "sqlserver".



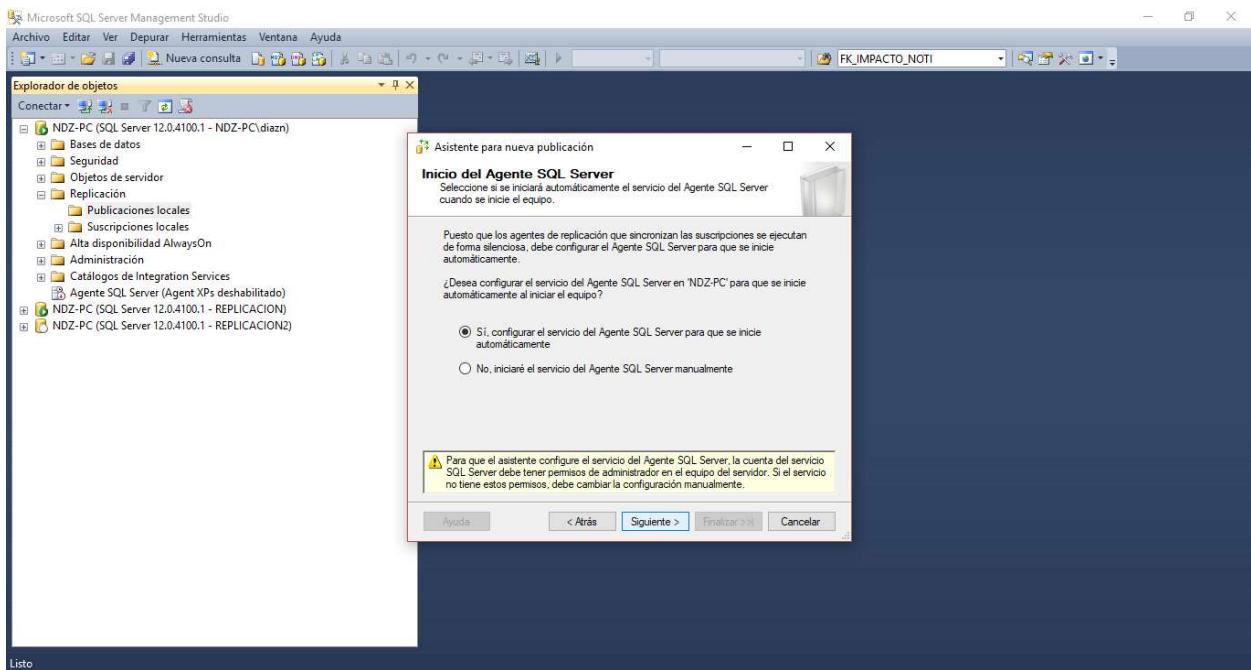
4. Procedemos a crear una nueva publicación.



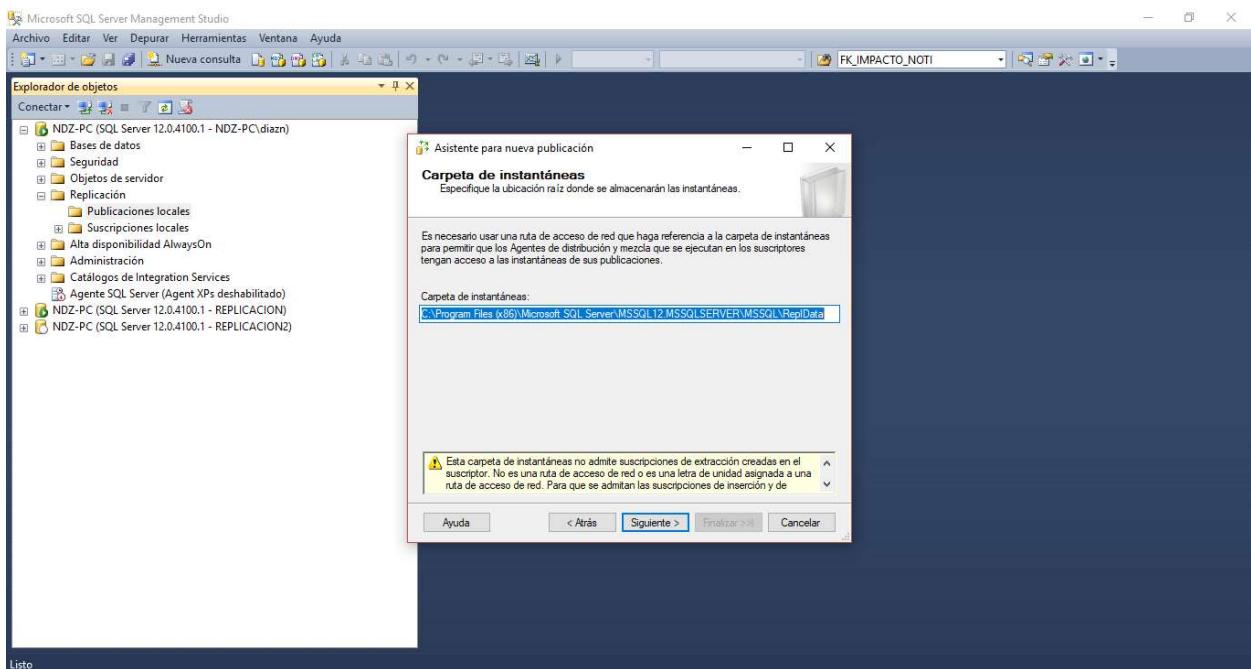
5. Elegimos el distribuidor.



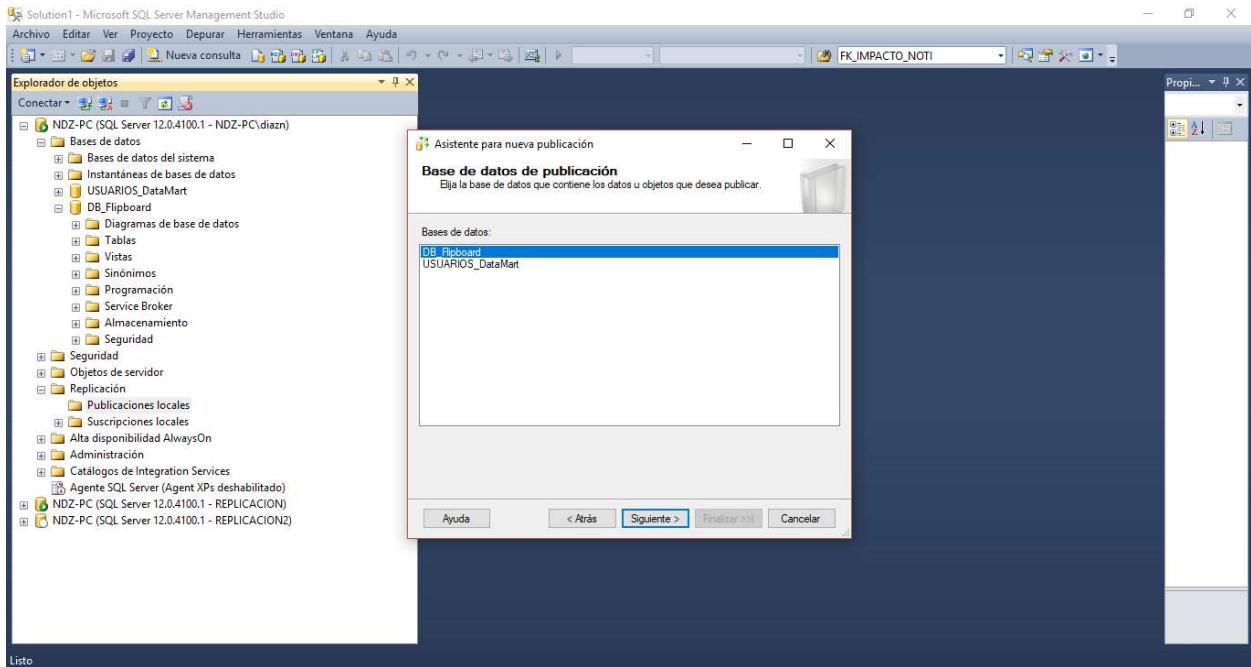
6. Configuramos el servicio de SQL Server para que se inicie automáticamente.



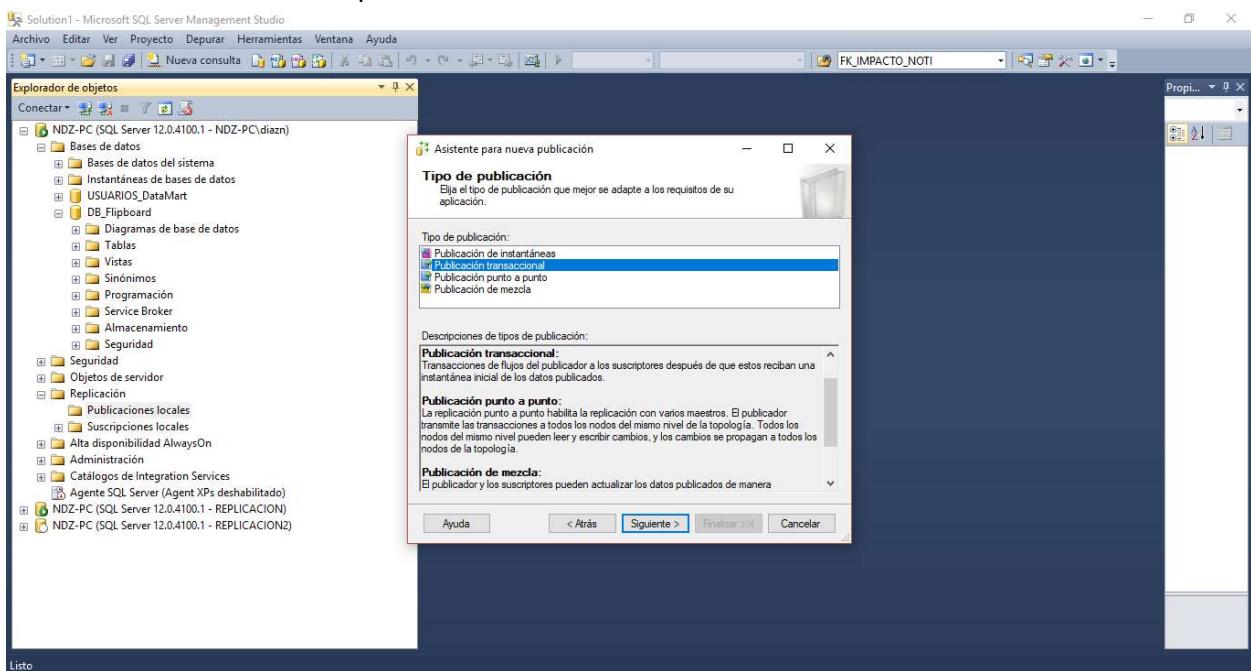
7. Elegimos una ruta para la carpeta de instantáneas.



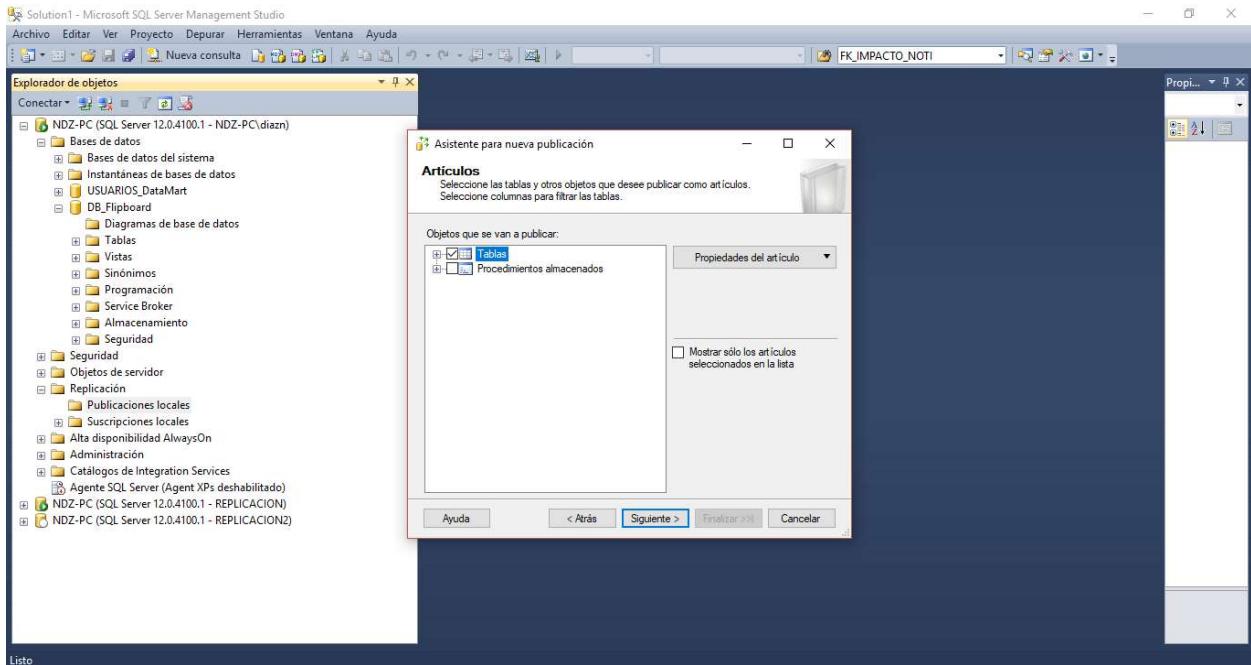
8. Elegimos la base de datos a replicar.



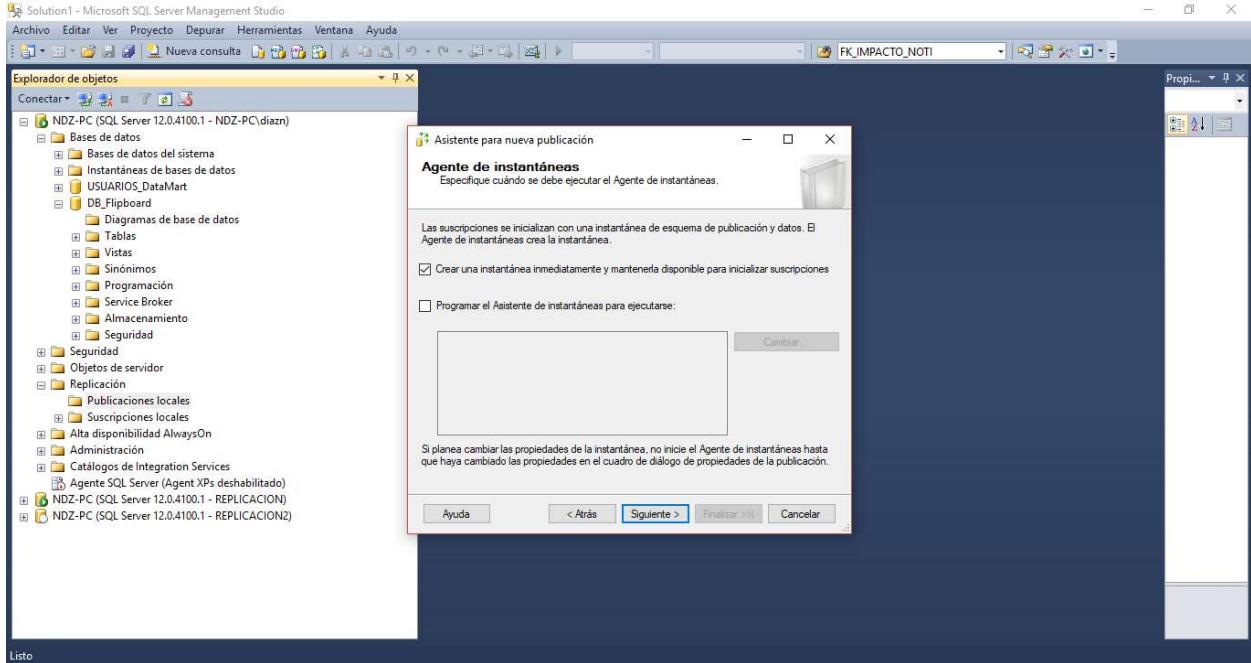
9. Seleccionamos una publicación transaccional.



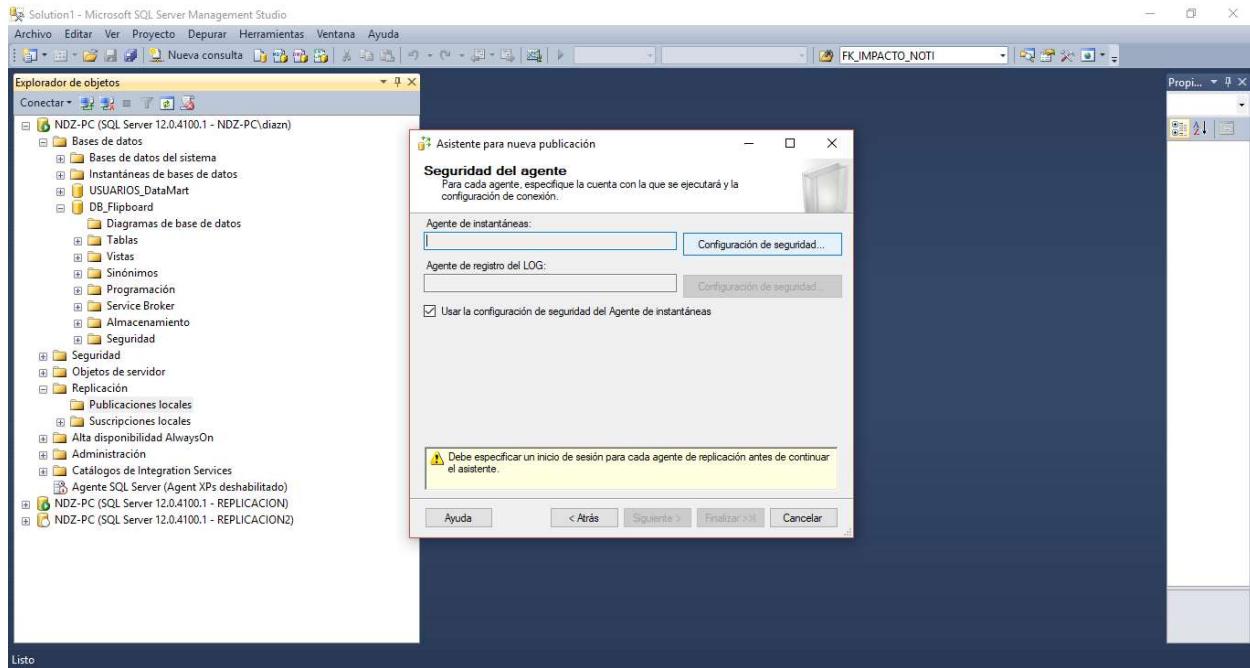
10. Seleccionamos todas las tablas para su replicación.



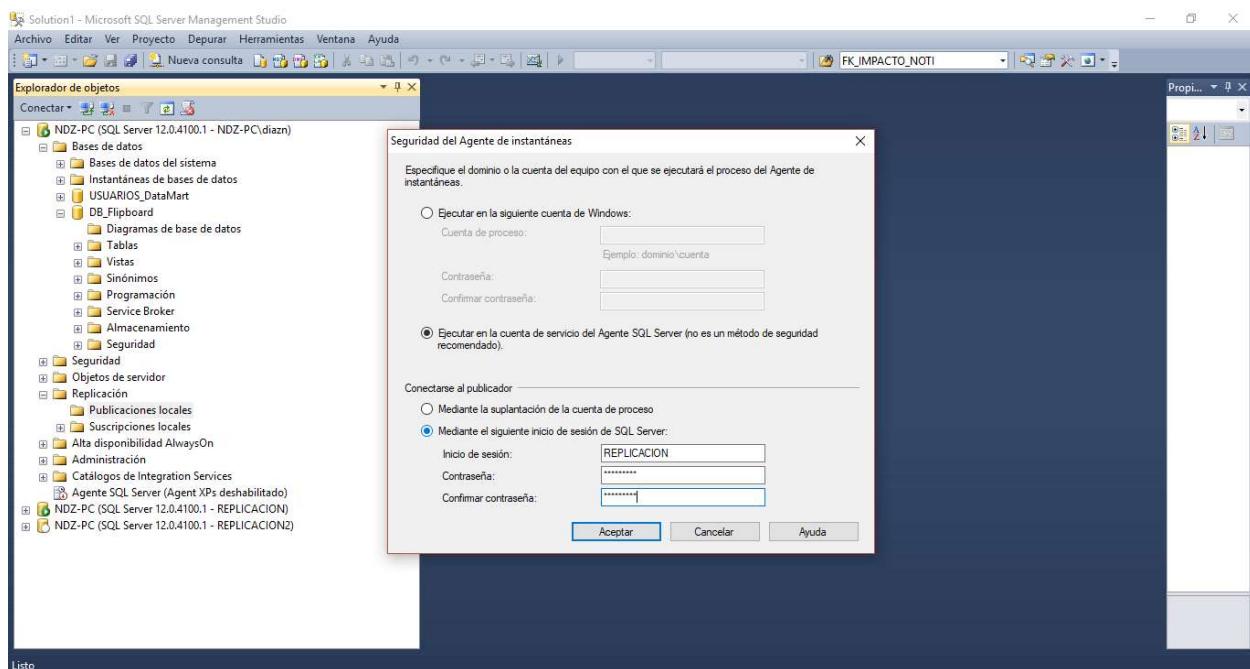
11. Nos aseguramos de que las instantáneas se creen inmediatamente.



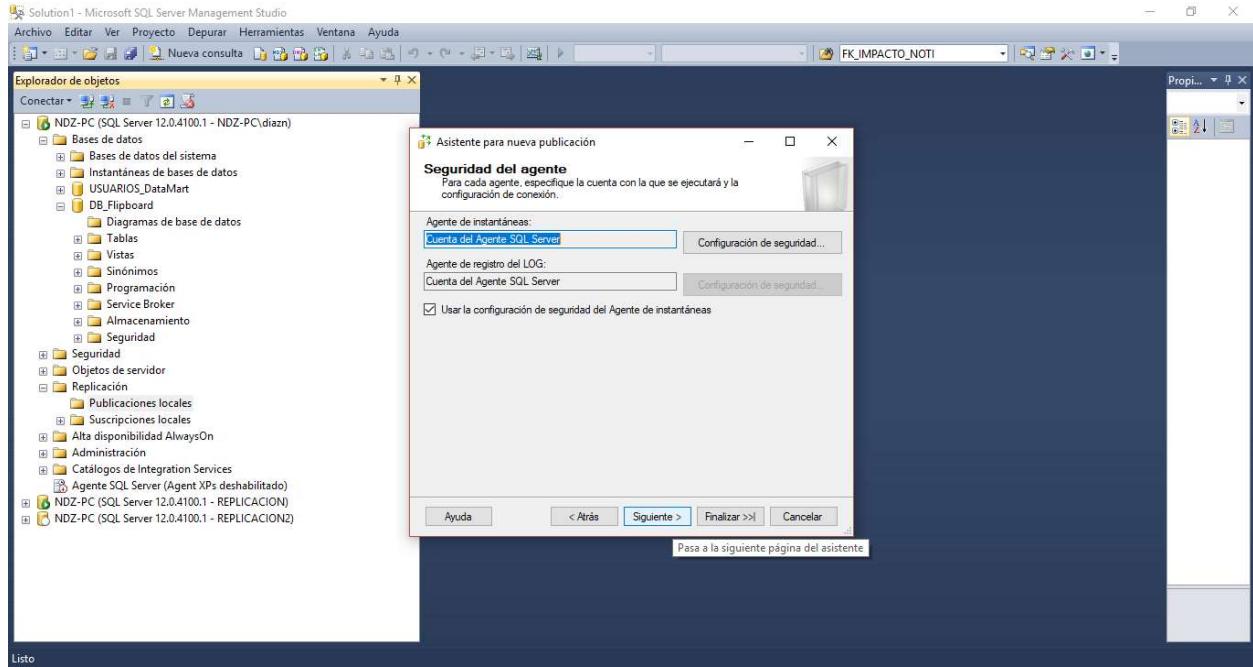
12. Configuramos la seguridad del agente.



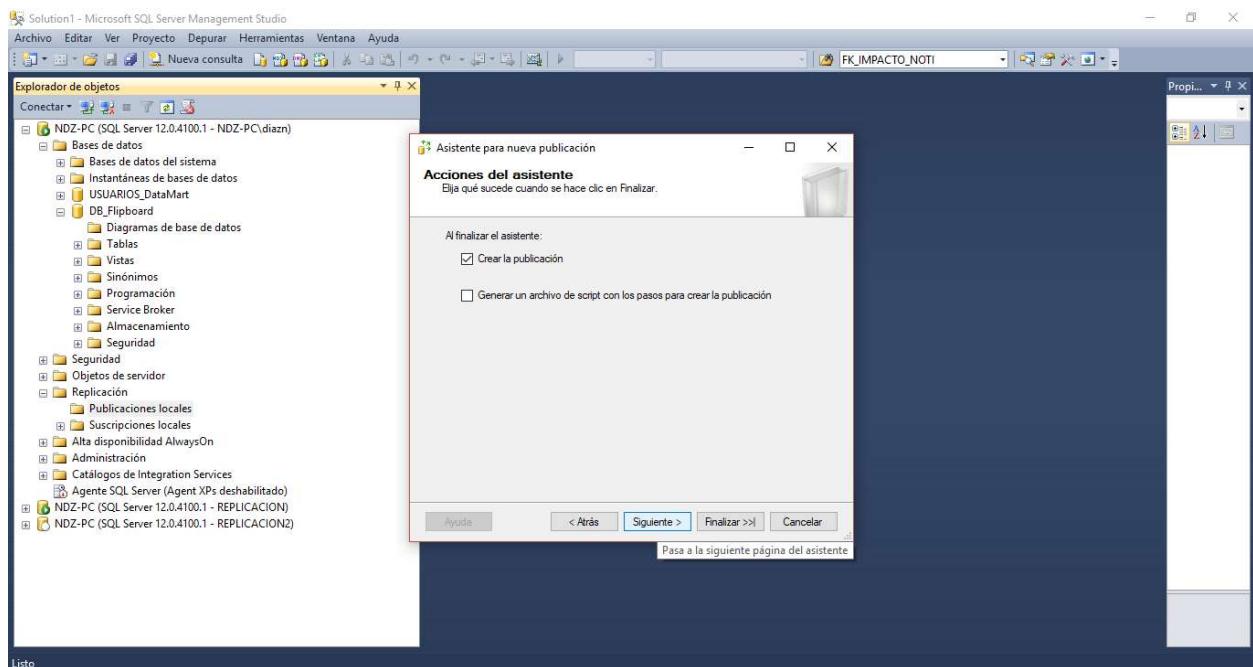
13. Configuramos para que se conecte a través del primer usuario creado.



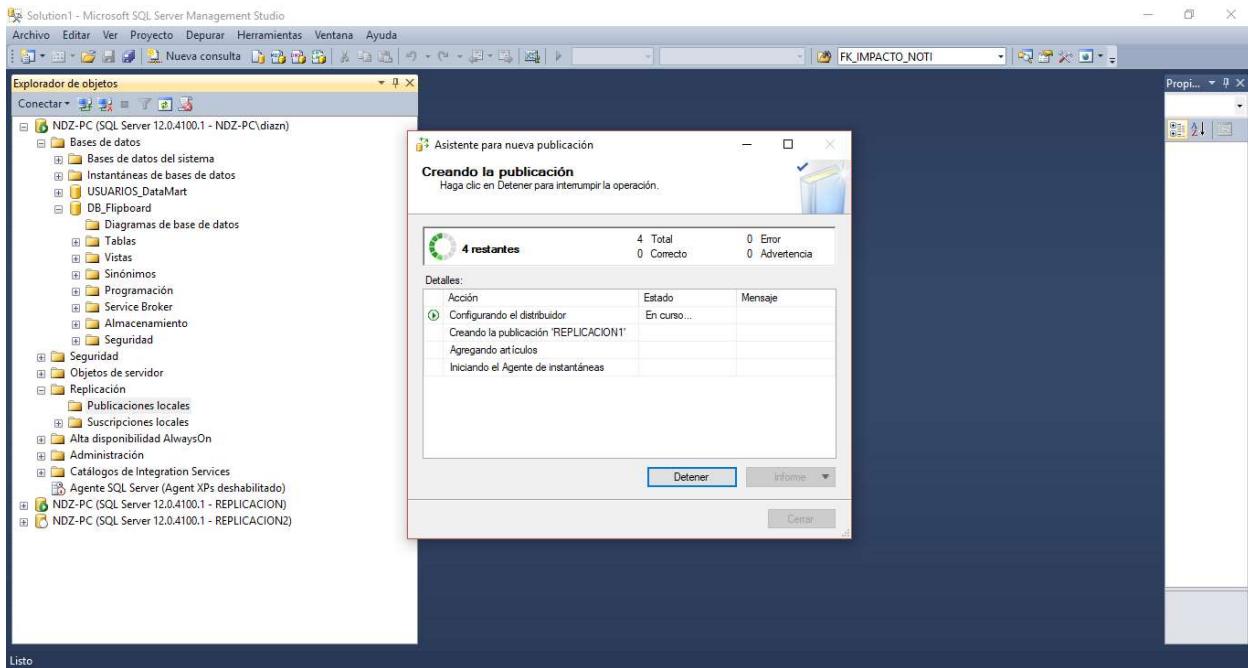
14. El agente esta configurado, damos a siguiente.



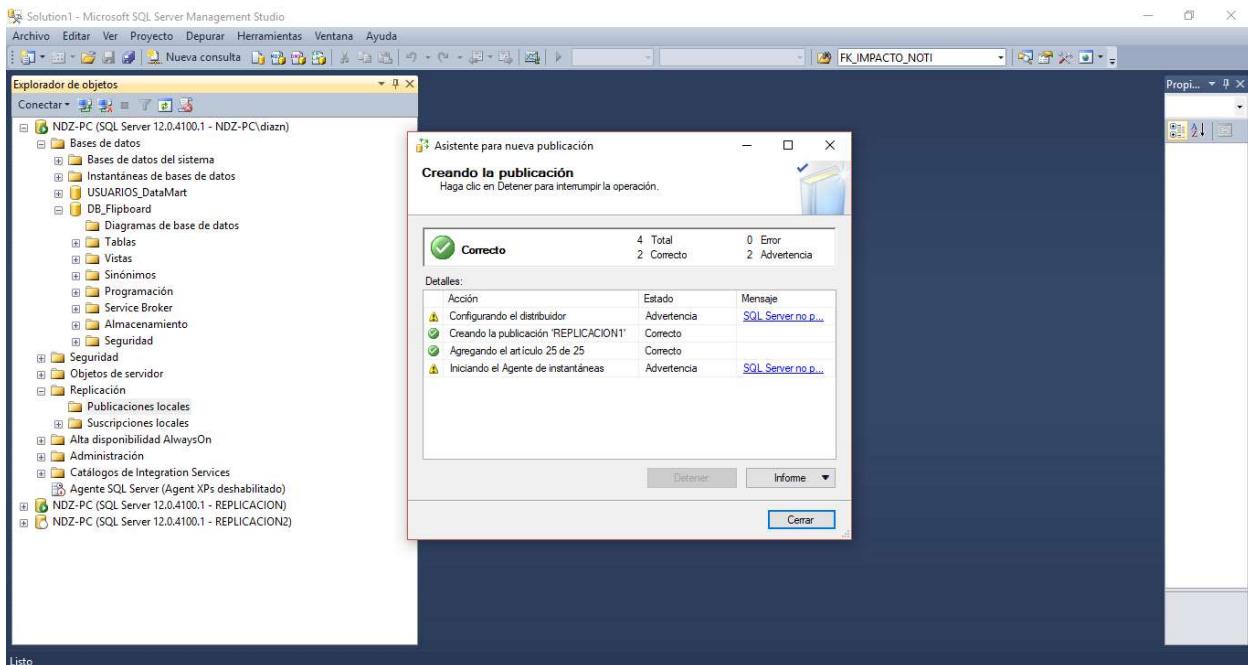
15. Empezamos los pasos para crear la publicación.



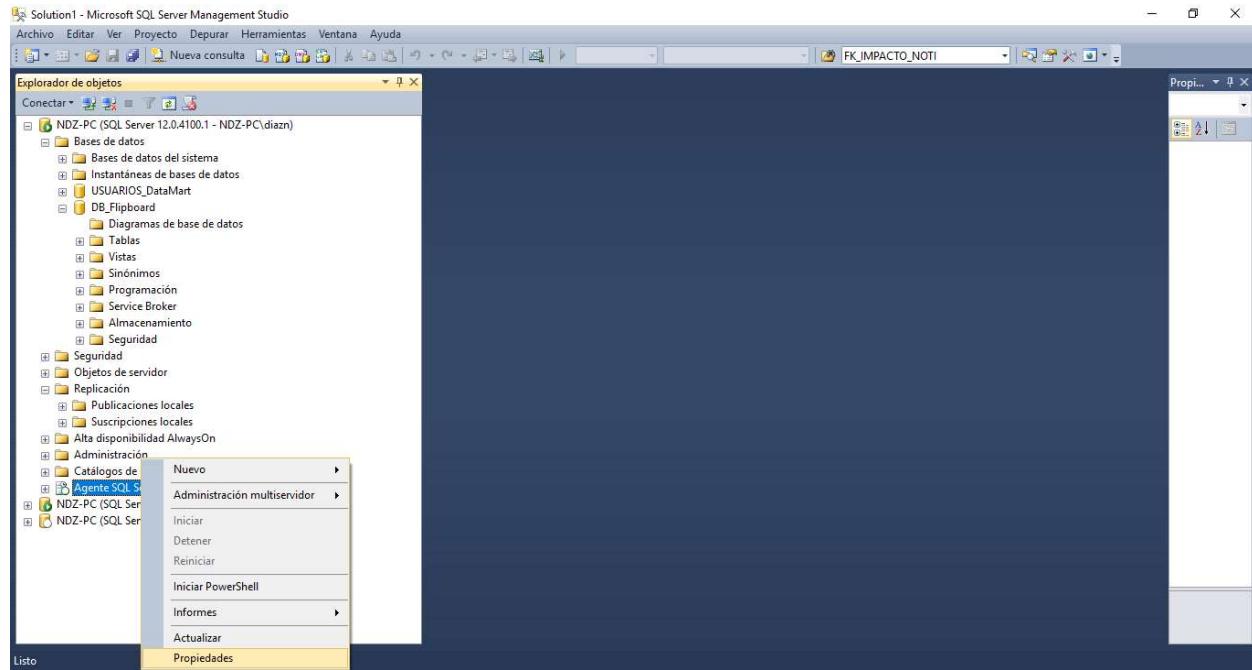
16. Elegimos el nombre de la publicación y dejamos que se cree.



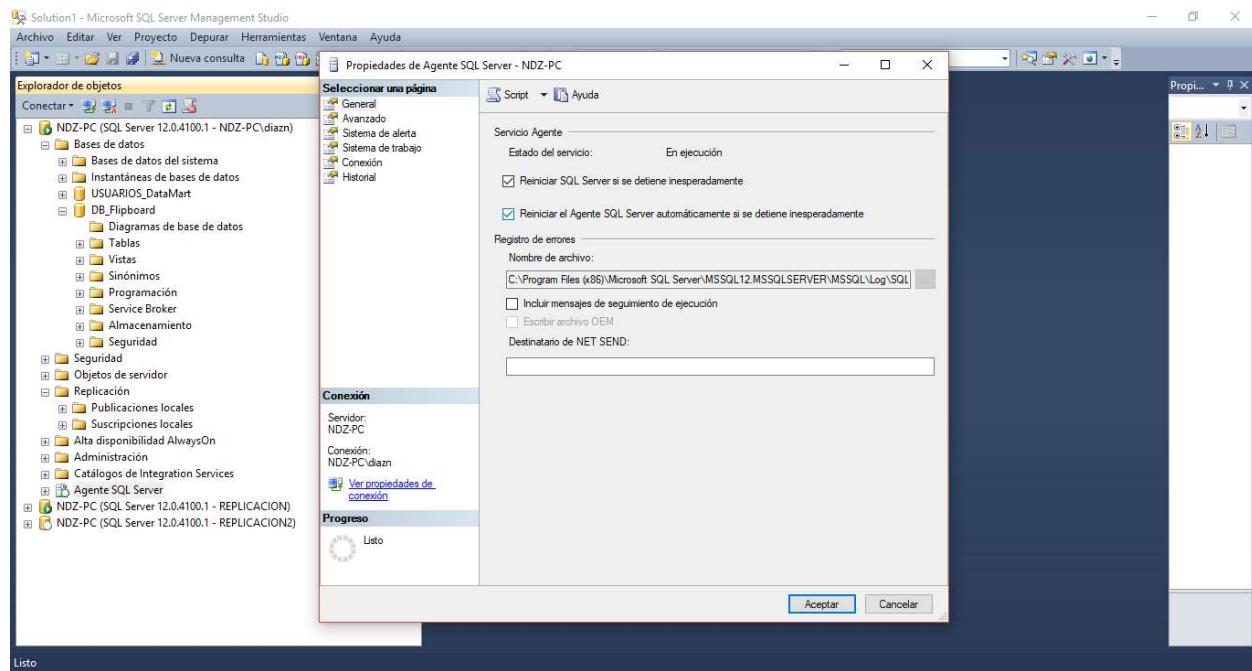
17. Una vez creado, si el agente no se configuro para iniciar automáticamente, deberemos hacerlo nosotros. Si el agente se configuro correctamente, saltar al paso 20.



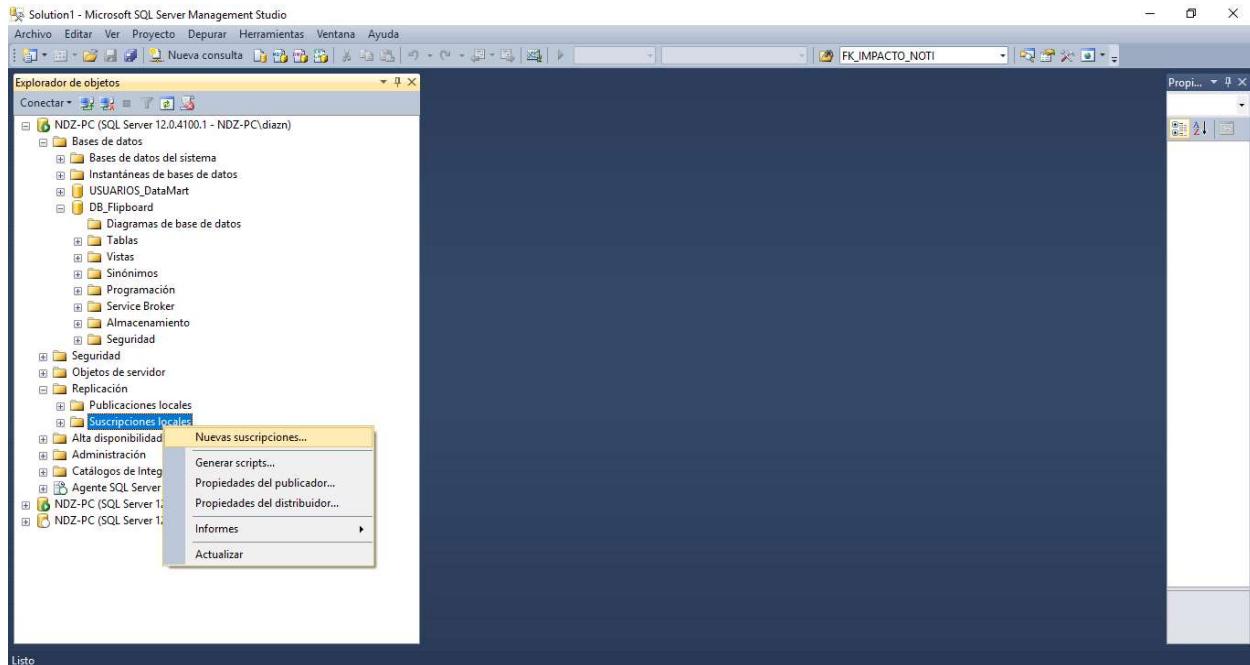
18. Clic derecho sobre el Agente SQL Server y seleccionamos propiedades.



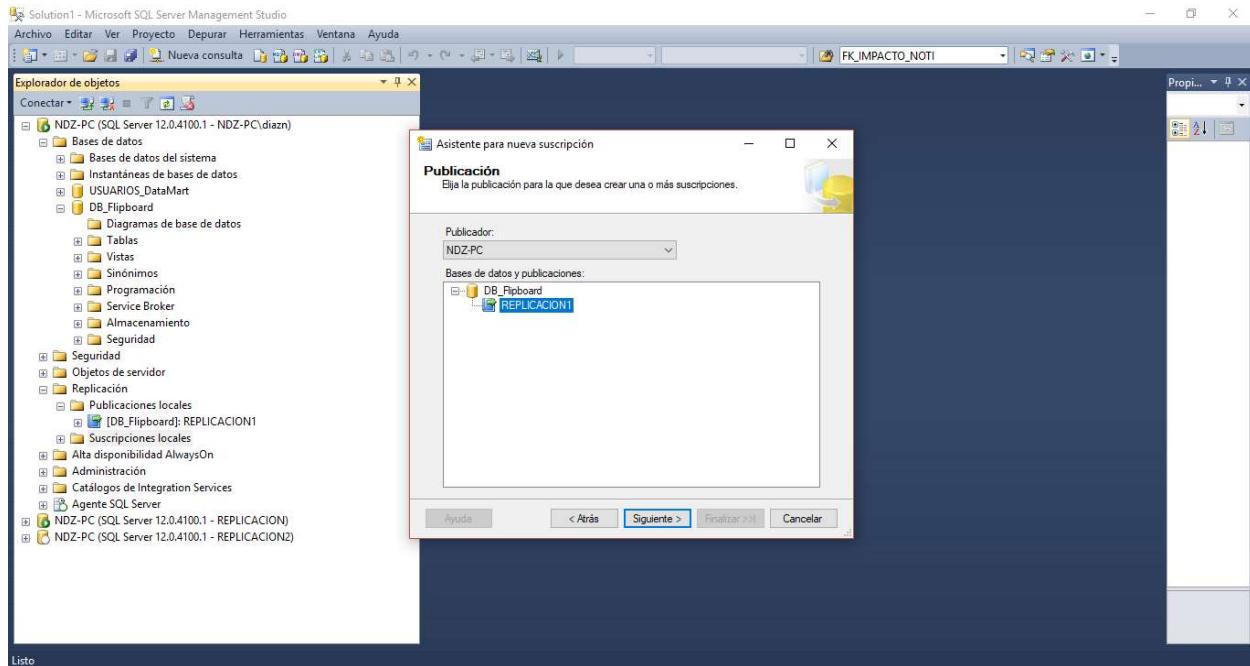
19. Hacemos que el Agente se reinicie automáticamente en caso de detenerse



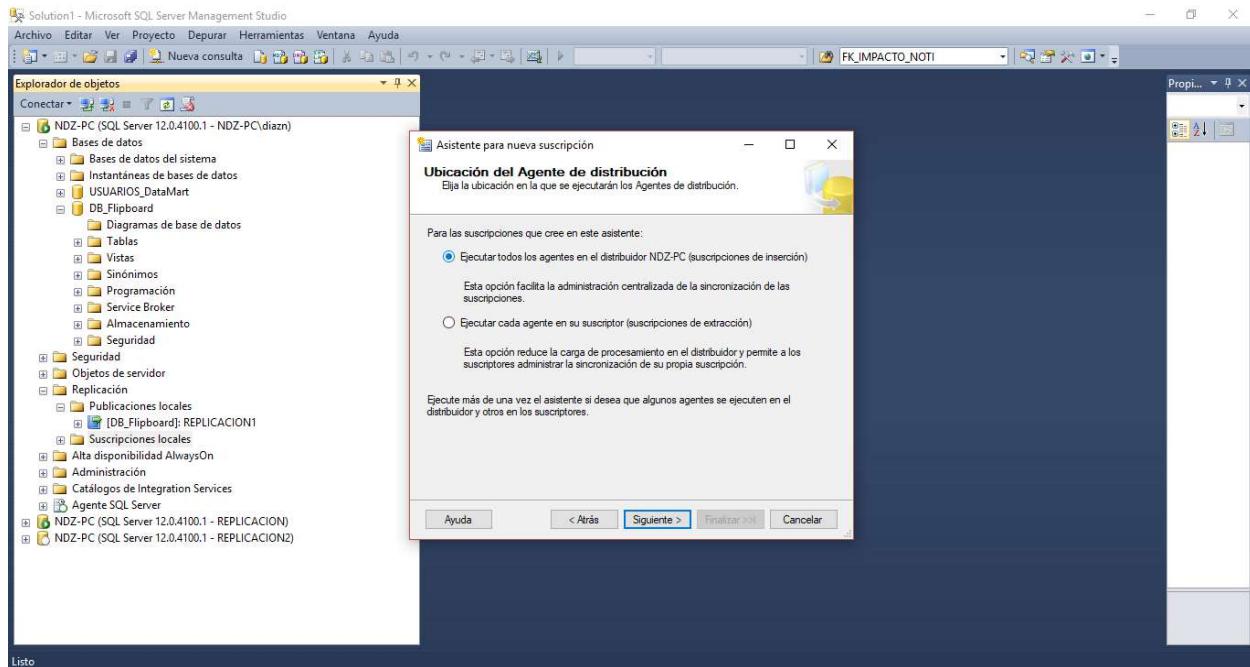
20. Procedemos a crear una nueva suscripción.



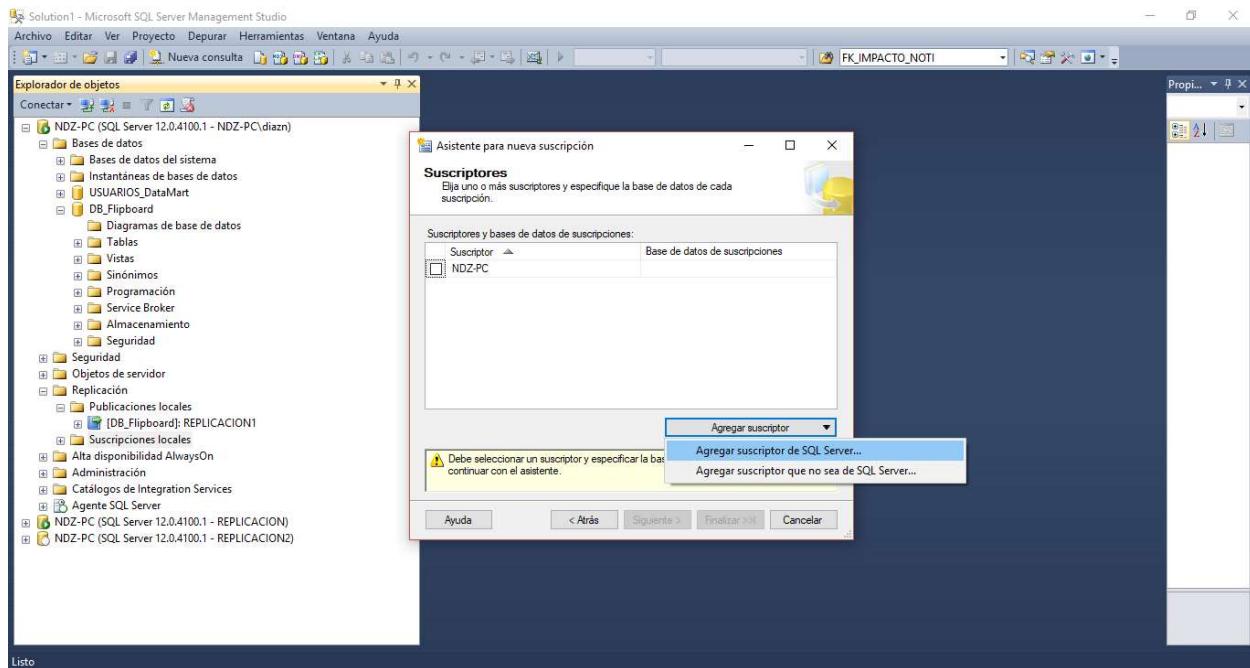
21. Seleccionamos la publicación antes creada.



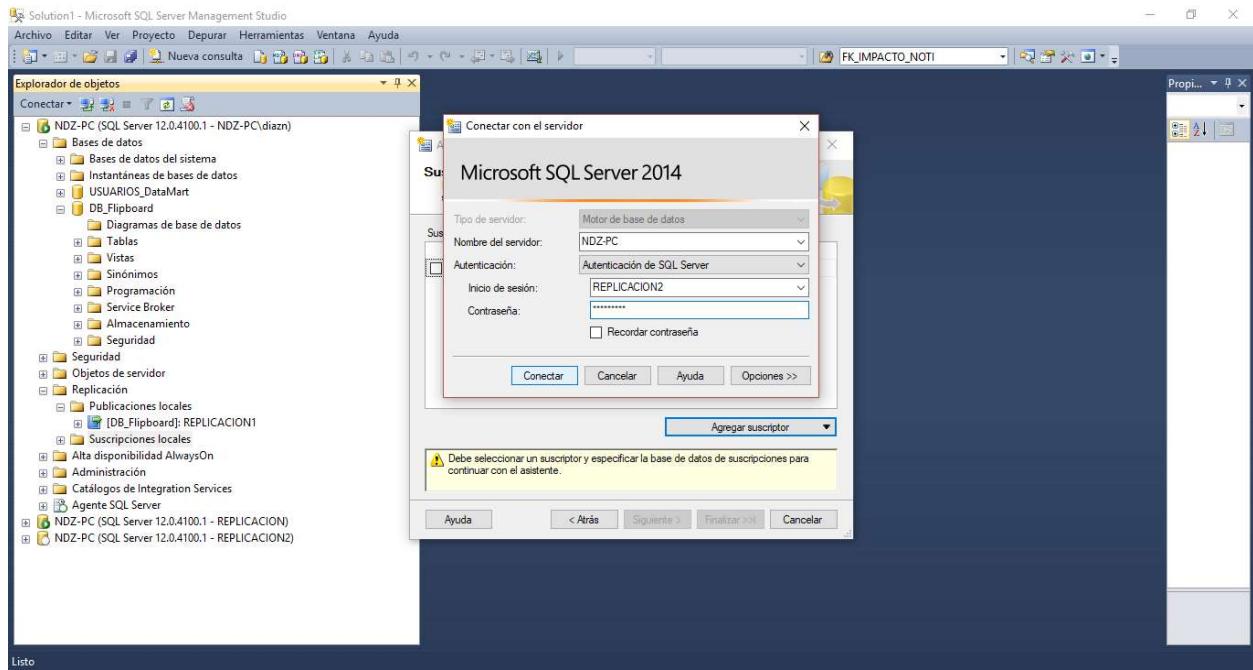
22. Ejecutamos todos los agentes en el distribuidor.



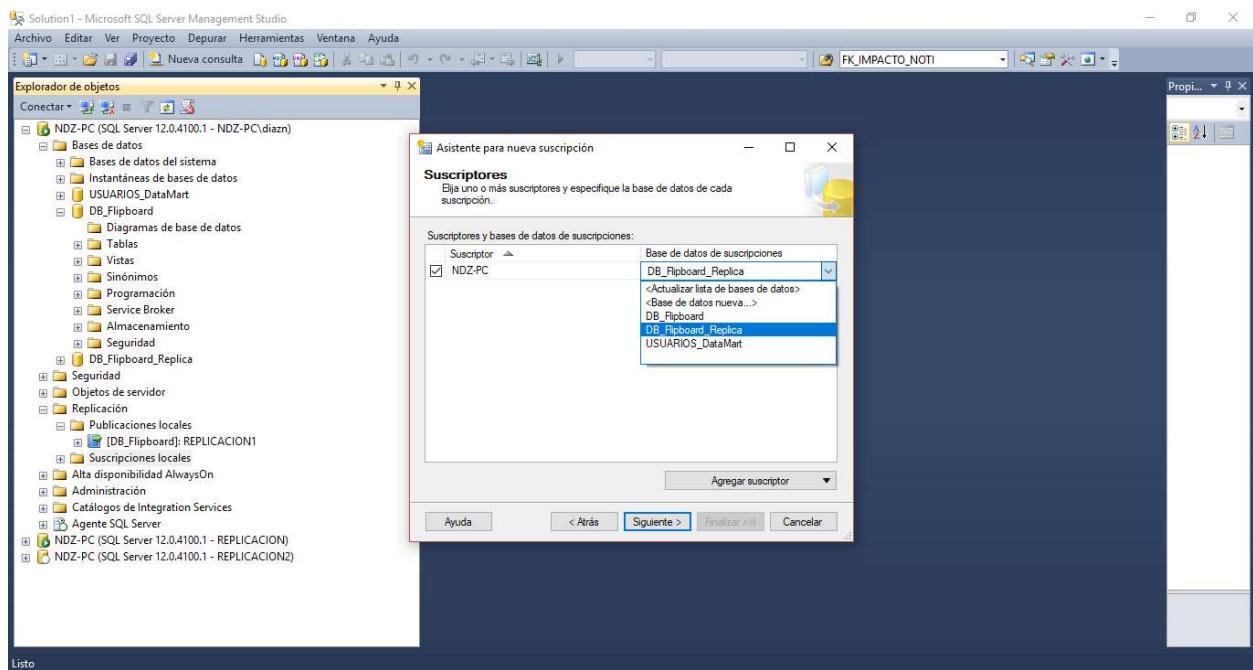
23. Agregamos un suscriptor de SQL Server



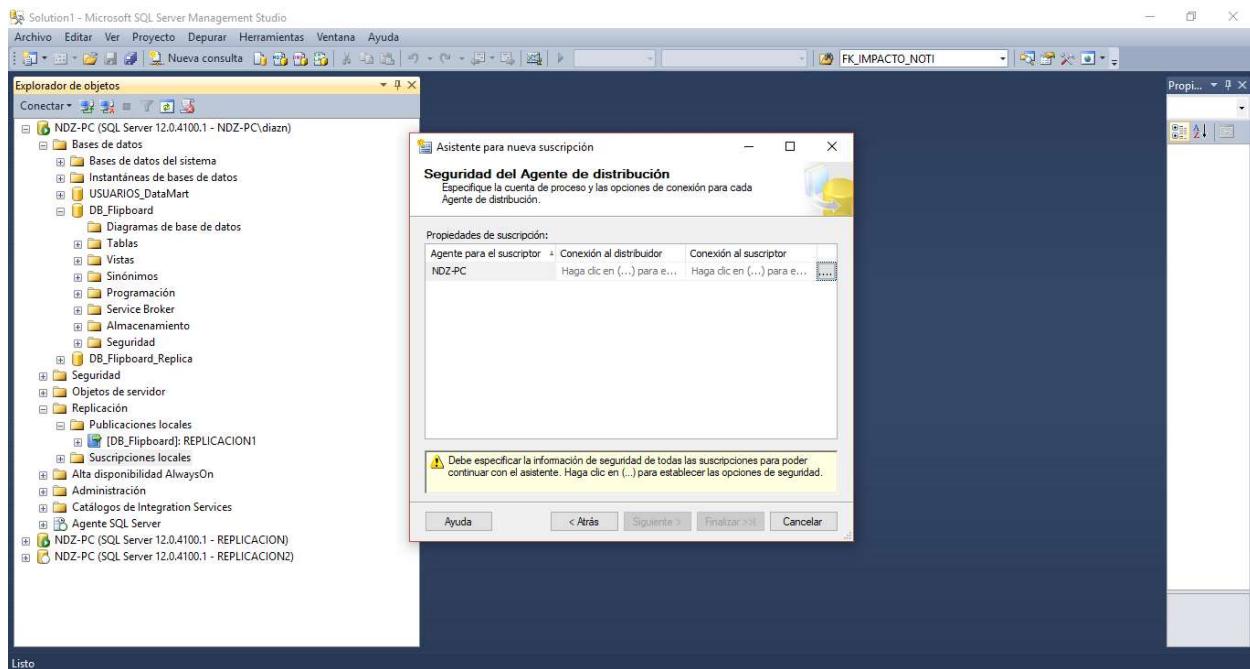
24. Conectamos con el segundo usuario que creamos.



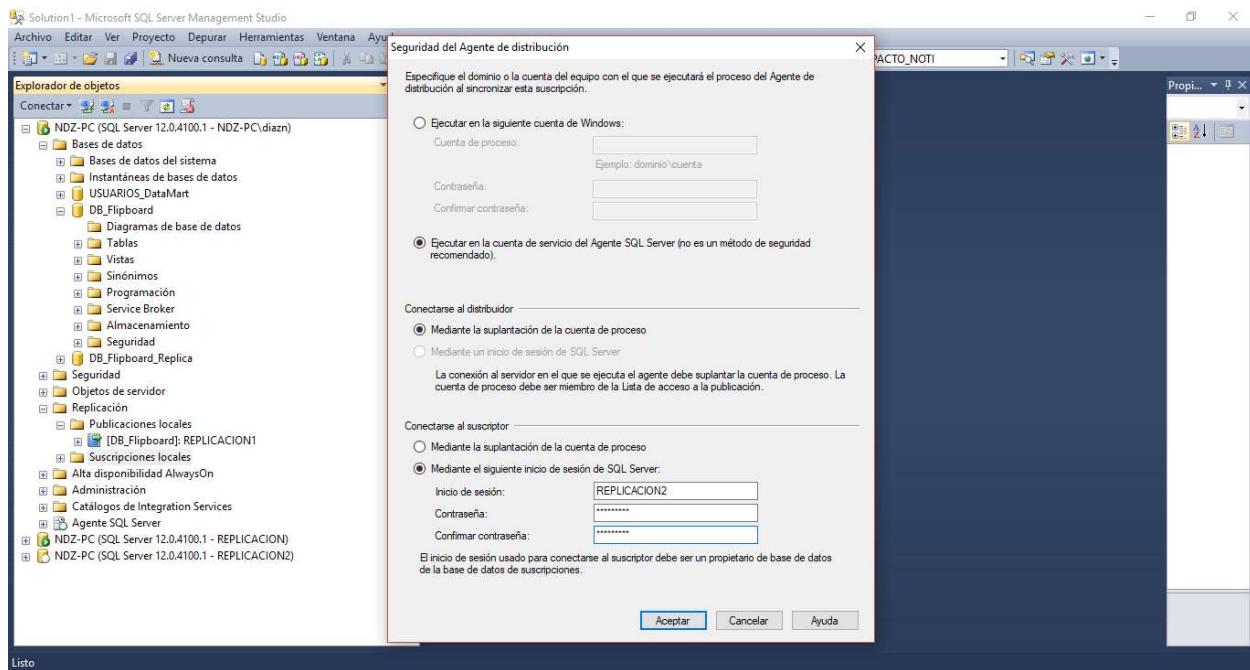
25. Seleccionamos la base de datos del suscriptor



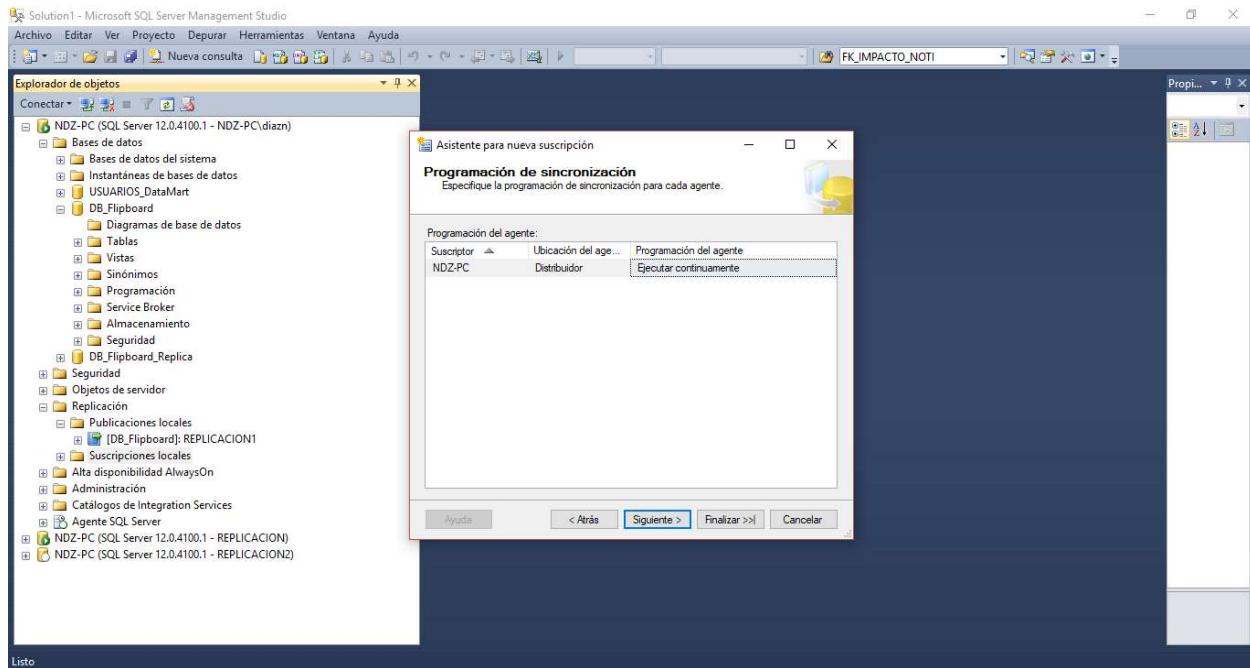
26. Configuramos el agente del suscriptor.



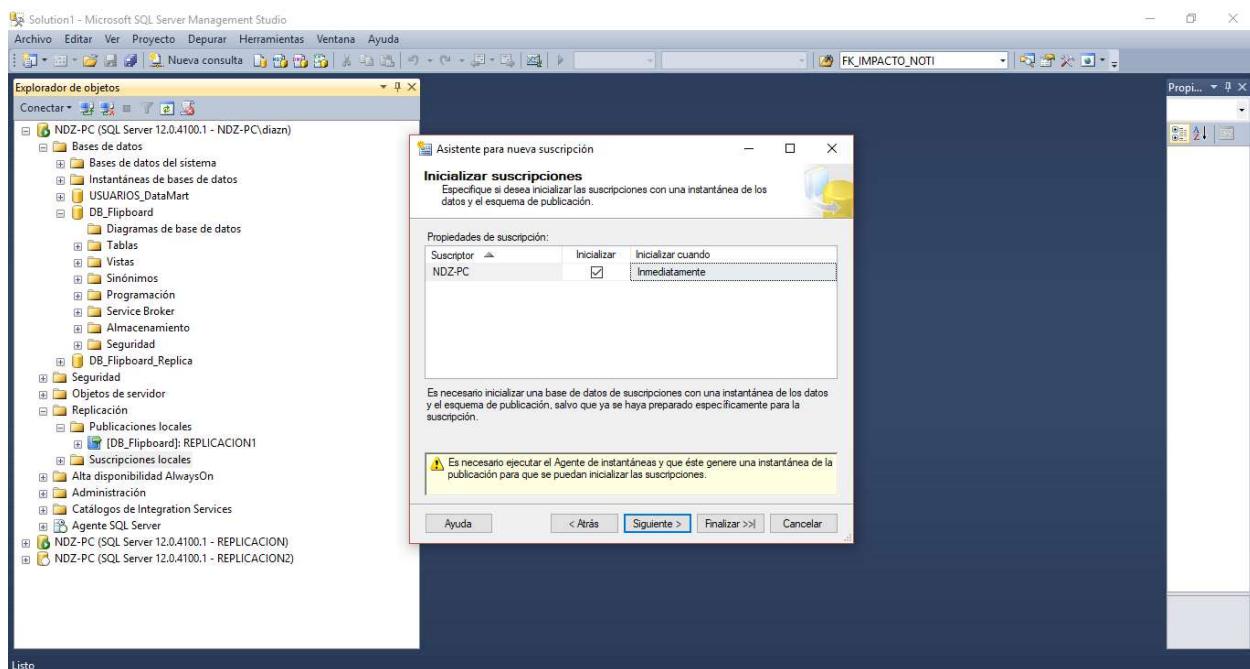
27. Configuramos como se muestra en la imagen.



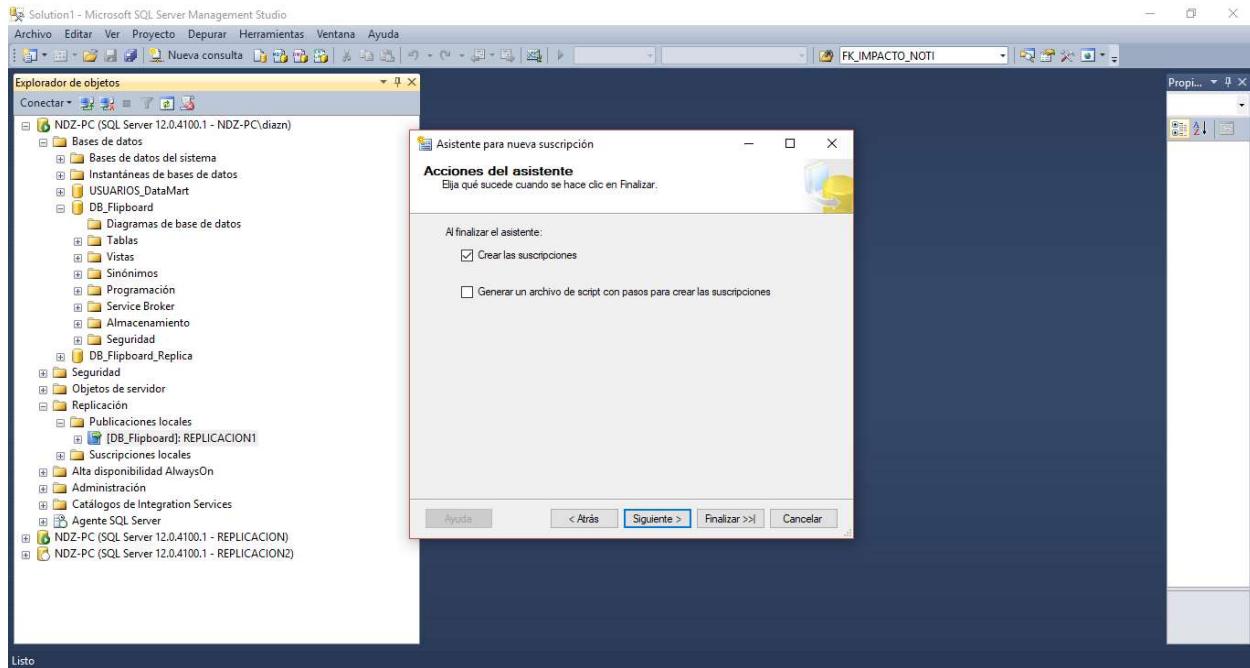
28. Seleccionamos que la sincronización se ejecute continuamente.



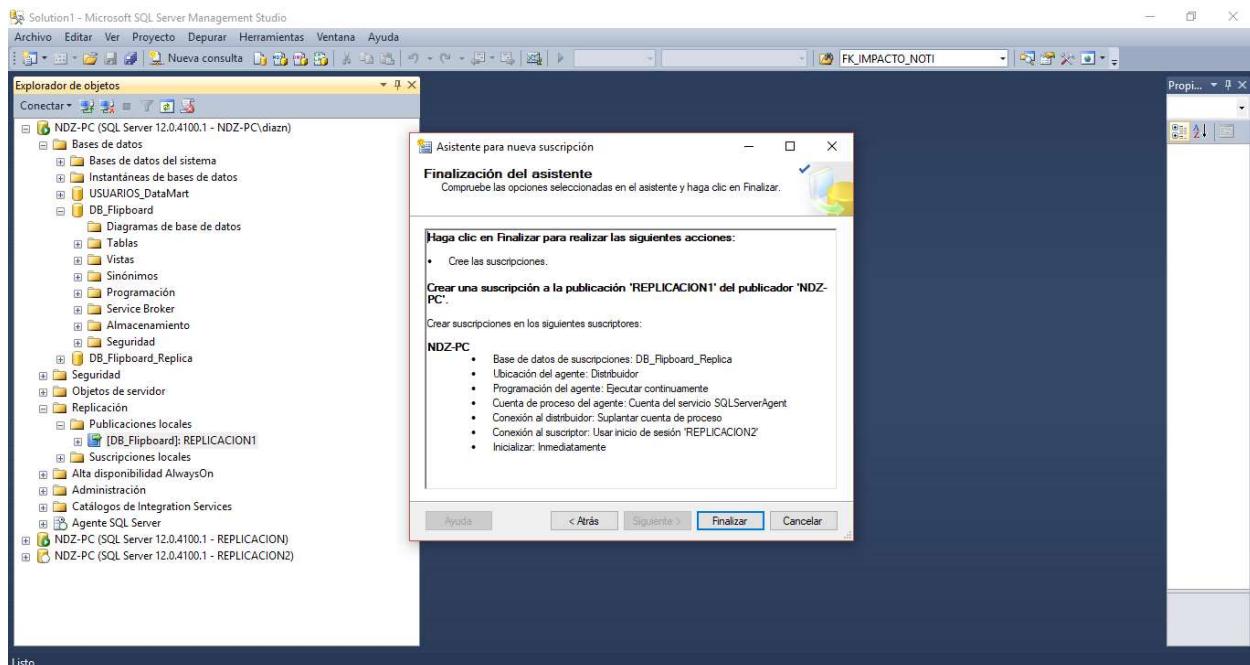
29. Inicializamos la suscripción inmediatamente.



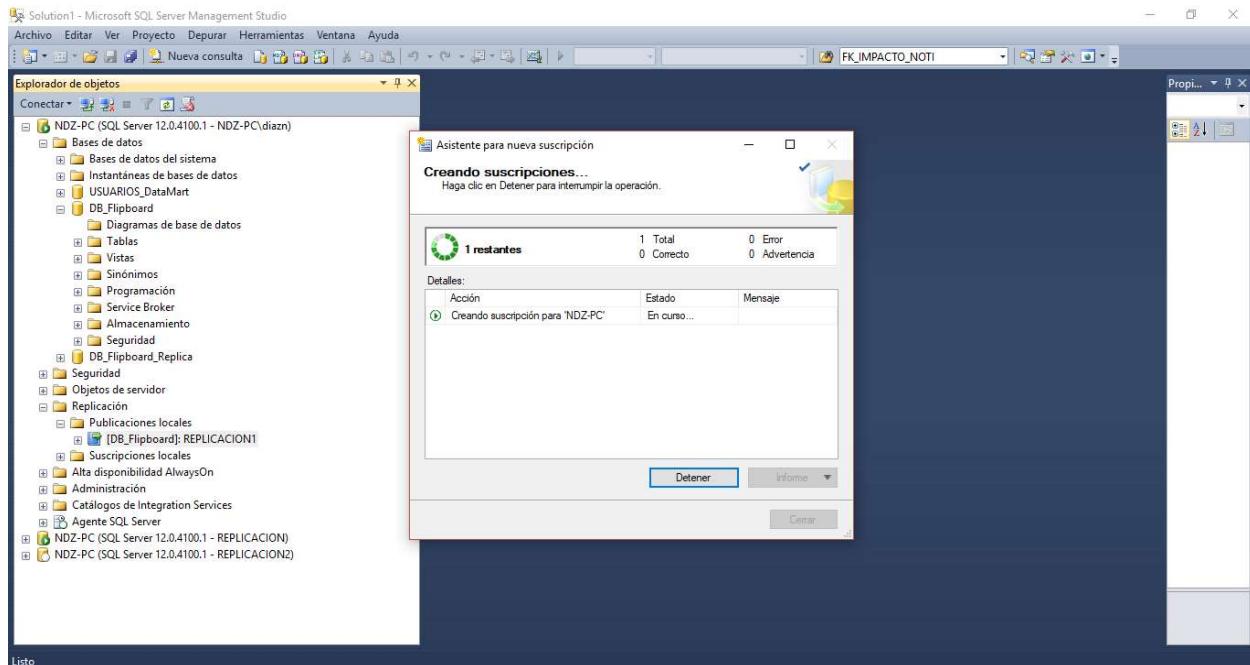
30. Elegimos que cree los suscriptores.



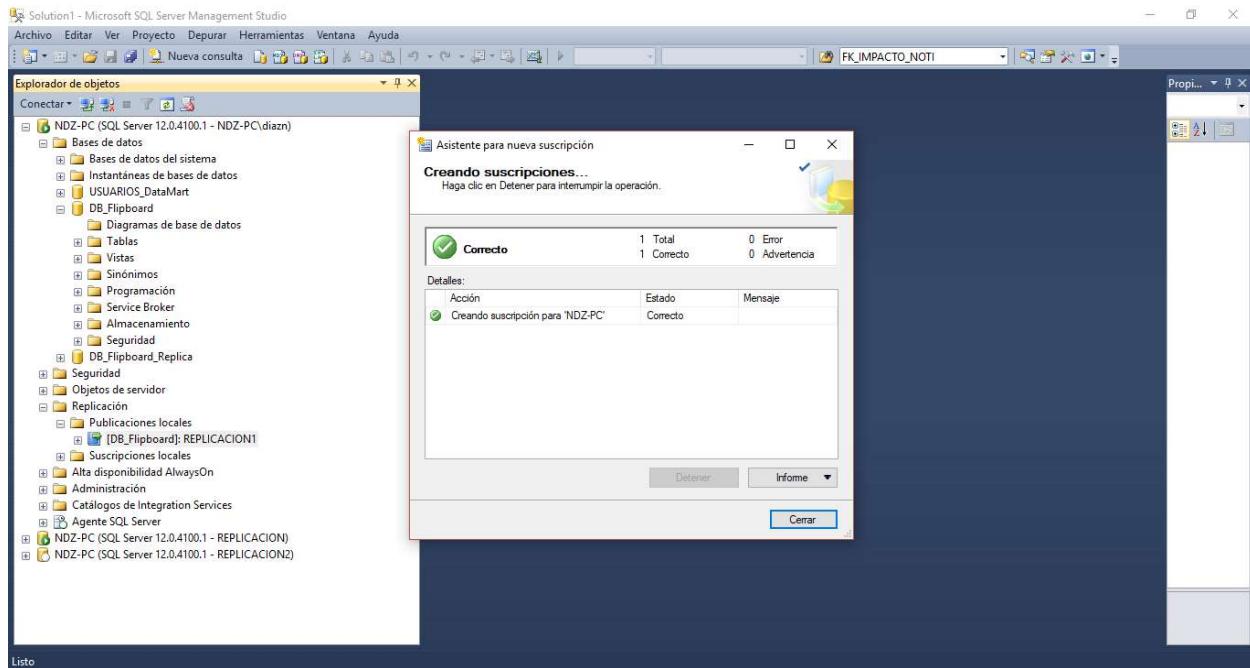
31. Finalizamos la creación del suscriptor.



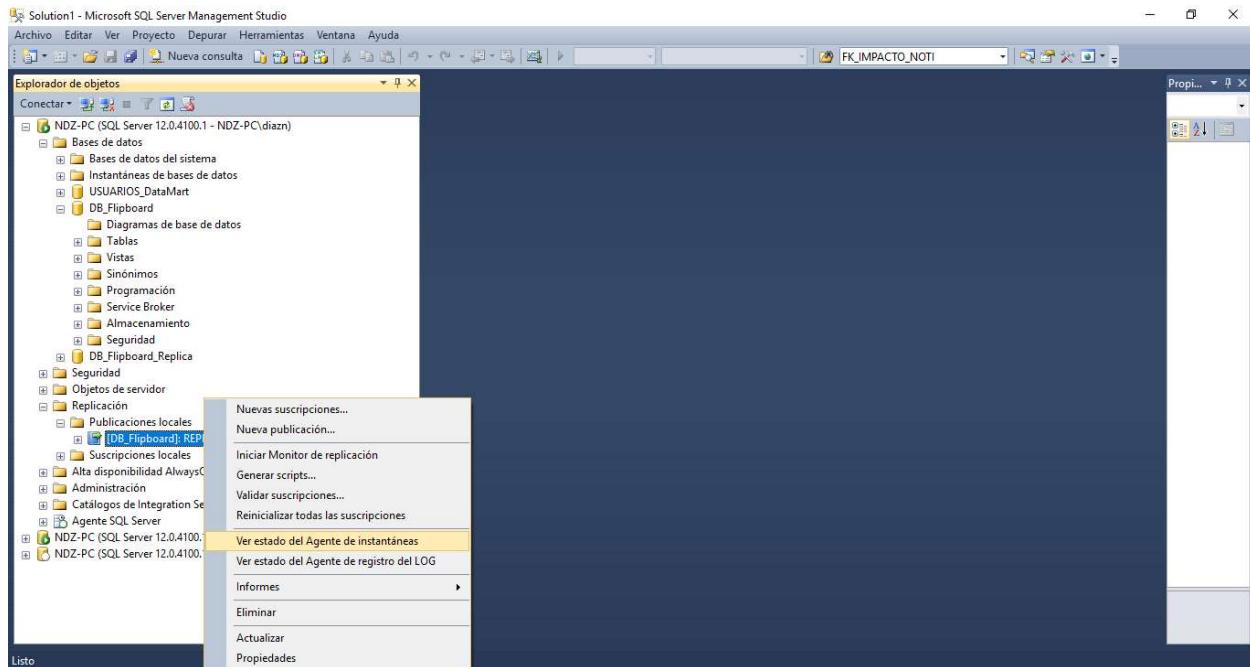
32. Se empieza a crear la suscripción.



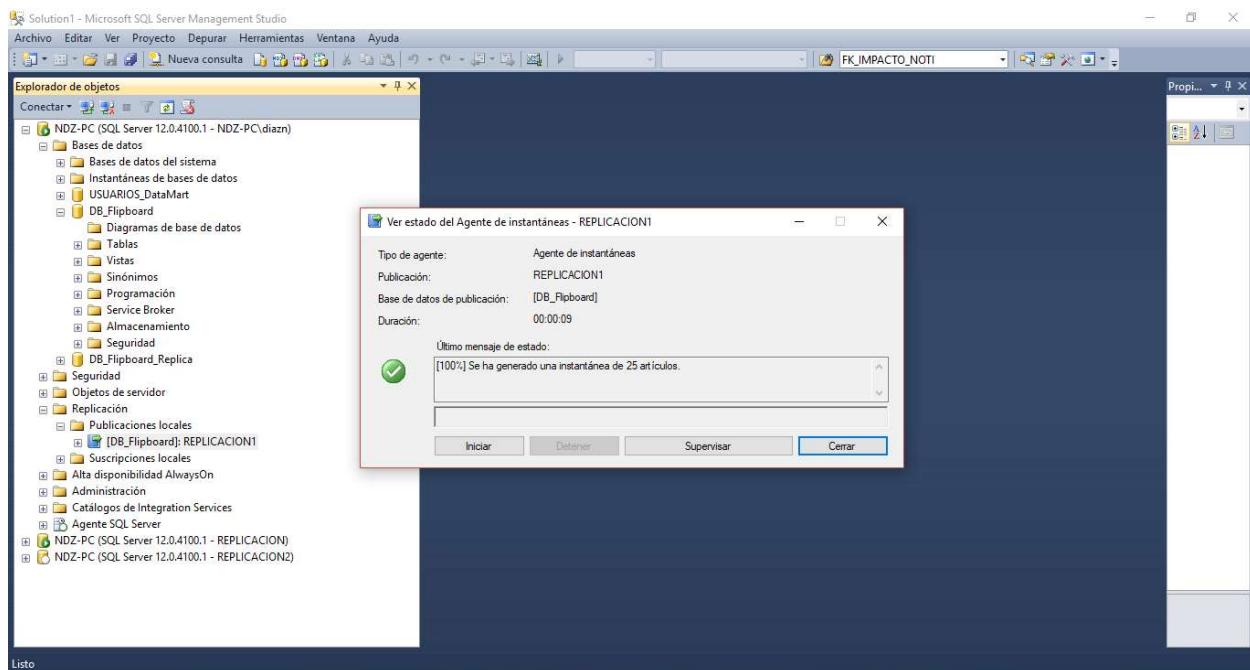
33. La suscripción se creó exitosamente.



34. Verificamos el estado del Agente de instantáneas.



35. Vemos que se ha generado correctamente una instantánea de 25 artículos (las tablas de la base de datos).



36. Todo está listo; para probar, se inserta un registro en cualquier tabla con el primer usuario y se verifica se refleje también en la base de datos del segundo usuario.