ANEXO 5 – INSTALACIÓN PENTAHO PDI – MIMIC-III

Detallamos a continuación el proceso de instalación de Pentaho PDI, la carga del catálogo y la instalación de la BBDD MIMIC-III.

Contenido

A5-1.	INSTALACIÓN PDI PENTAHO	3
A5-2.	INSTALACIÓN MIMIC-III)

A5-1. INSTALACIÓN PDI PENTAHO

El primer paso es descargar el módulo PDI de la web de Pentaho <u>http://community.pentaho.com/</u>,utilizamos la versión community, la disponible a fecha mayo 2017 es la 7.0 y se encuentra en el apartado *Downloads* \rightarrow *Data integration*.

Es necesario disponer de java instalado, 1.6 o superior, versión de 64bits, antes de proseguir la instalación. La instalación se ha realizado en un equipo linux, en un equipo windows se realiza de forma similar.

Guardamos el fichero *pdi-ce-7.0.0.0-25.zip* en el directorio de instalación y lo descomprimimos.

revirier@hyperion:/opt/PDI\$ sudo mv ~/Downloads/pdi-ce-7.0.0.0-25.zip . revirier@hyperion:/opt/PDI\$ sudo unzip pdi-ce-7.0.0.0-25.zip . revirier@hyperion:/opt/PDI\$ ls -1 data-integration revirier@hyperion:/opt/PDI\$ ls -l data-integration/ total 9528-rw-r--r--1 root root551 nov5 2016 Carte.bat-rwxr-xr-x1 root root405 nov5 2016 carte.shdrwxr-xr-x2 root root4096 nov5 2016 Data Integration.appdrwxr-xr-x3 root root4096 nov5 2016 Data Service JDBC Driverdrwxr-xr-x3 root root4096 nov5 2016 Carte.shdrwxr-xr-x2 root root4096 nov5 2016 Data Service JDBC Driverdrwxr-xr-x3 root root4096 nov5 2016 Data Service JDBC Driverdrwxr-xr-x1 root root138 nov5 2016 Encr.bat-rw-r--r--1 root root131 nov5 2016 Import.bat-rw-r--r--1 root root2354 nov5 2016 import-rules.xml-rwxr-xr-x1 root root275 nov5 2016 Kitchen.bat-rwxr-xr-x1 root root354 nov5 2016 kitchen.shdrwxr-xr-x2 root root4096 nov5 2016 kitchen.sh total 9528 drwxr-xr-x 2 root root 20480 nov 5 2016 lib drwxr-xr-x 6 root root 4096 nov 5 2016 libswt -rw-r--r-- 1 root root 13366 nov 5 2016 LICENSE.txt -rw-r--r-- 1 root root 172 nov 5 2016 Pan.bat -rwxr-xr-x 1 root root 320 nov 5 2016 pan.sh -rw-r--r-- 1 root root 9181335 nov 5 2016

 PentahoDataIntegration_OSS_Licenses.html

 drwxr-xr-x 27 root root
 4096 nov 5
 2016 plugins

 -rw-r-r-r
 1 root root
 213 nov 5
 2016 purge-utility.bat

 -rwxr-xr-x 1
 root root
 347 nov 5
 2016 purge-utility.bat

 -rwxr-xr-x 2
 root root
 4096 nov 5
 2016 purge-utility.sh

 drwxr-xr-x 2
 root root
 1312 nov 5
 2016 README.txt

 -rw-r--r- 1 root root
 522 nov 5
 2016 runSamples.bat

 -rwxr-xr-x 1
 root root
 305 nov 5
 2016 samples

 -rwxr-xr-x 5
 root root
 4096 nov 5
 2016 samples

 -rwxr-xr-x 1
 root root
 305 nov 5
 2016 set-pentaho-env.bat

 -rwxr-xr-x 1
 root root
 4098 nov 5
 2016 set-pentaho-env.bat

 -rwxr-xr-x 1
 root root
 3710 nov 5
 2016 simple-jndi

 -rwxr-xr-x 1
 root root
 3857 nov 5
 2016 spoon.bat

 -rwxr-xr-x 1
 root root
 98 nov 5
 2016 SpoonConsole.bat

 -rwxr-xr-x 1
 root root
 1271 nov 5
 2016 SpoonDebug.bat

 -rw-r--r- 1 root root
 1051 nov 5
 2016 SpoonDebug.sh
 PentahoDataIntegration OSS Licenses.html -rw-r--r-- 1 root root 370070 nov 5 2016 spoon.ico -rw-r--r-- 1 root root 1345 nov 5 2016 spoon.png

```
-rwxr-xr-x1 root root6307 nov5 2016 spoon.shdrwxr-xr-x5 root root4096 nov5 2016 systemdrwxr-xr-x3 root root4096 nov5 2016 ui-rwxr-xr-x1 root root1628 nov5 2016 yarn.sh
```

Antes de ejecutarlo, modificamos la memoria disponible editando el fichero spoon.sh y ampliando la memoria en la sección de JAVA_OPTIONS:

```
if [ -z "$PENTAHO_DI_JAVA_OPTIONS" ]; then
        PENTAHO_DI_JAVA_OPTIONS="-Xms1024m -Xmx4096m -XX:MaxPermSize=256m"
fi
```

Y creamos las BBDD necesarias para el repositorio y las tablas intermedias, en este caso las creamos en PosgreSQL, pero se pueden crear en cualquiera de los gestores de BBDD soportados por Pentaho.



Una vez disponibles las BBDD necesarias: mimicat02 (catálogo), mimic-iii y mimicsel ejecutamos el script *spoon.sh*. En la pantalla principal de Pentaho se debe seleccionar la opción de *connect* \rightarrow *repository manager*



Y elegir la opción other repositories -> Database Repository

×	Spoon - Welcome!				
File Edit View Action Tools Help					
		Connect •			
🖻 View 🖉 Design 🛛 🔀 Welco	Repository Manager				
Explorer Q 2 = = • • •	Other Repositories Not supported or recommended for production use	×			
Wel	Database Repository Unsupported ETL storage that uses a database	Many Reasons to Get			
	File Repository Unsupported ETL storage that uses your local file system	Enterprise Edition			
	Pentaho Repository				
	Take your ETL, Analytics & Big Data to the	e Next Level			
	When the define one in the second second second for the second second second second second second second second	۲			

Creamos una conexión de BBDD nueva:



*		Spoon - Welcome!	
File Edit View Action Too	ls Help		
			Connect -
View Ø Design	🔀 Welcc 🔛	Repository Manager	
Explorer 🔍 🖌 🖫		Connection Details	
zdoL 🗂	Pe	Display Name	
		mimiccat02	,
	Wel	Database Connection	Many Reasons to Get
		mimiccat02 >	Enterprise Edition
		Description	
		Database repository	
		Launch connection on startup	TRACTORY CO.
	Pel ? Hel	p Back Finish	
	•	Take your ETL, Analytics & Big Data to the N	Next Level

Y nos conectamos a ella:



La conexión a los catálogos se realiza con el usuario *admin /* clave *admin* por defecto:

*	Spoon - Welcome!	
File Edit View Action Too	s Help	
		Connect -
View Ø Design	🐮 welcc 😤 Repository Manager 🛛 🕅	
Explorer 🔍 🖌 🖫	• •	
Transformations		
Jobs	Vel Wel Connect to mimiccat02 User name admin Password Image: Connect to Connect	Many Reasons to Get Enterprise Edition
	Take your ETL, Analytics & Big Data to the Nex	t Level
	Mark & Astronom Incolor and the Astronom Incolor and Astronom Incolor and Astronom Incolor and Astronomy Incol and Astronomy Incol a	

Una vez conectados al repositorio debemos importar el fichero XML que contiene la exportación del catálogo desarrollado en este trabajo, en el menú **Tools**-> **Import Repository** elegimos el fichero correspondiente:

*		Spoon - Welcome!				
File Edit View Action Tools Help						
				admin mimie	cat02 •	
View Ø Design	🔀 W 😹	Select an XML file to import from				
Explorer Q 4 8=	A Sevirier	Master_ICM TFM Pentaho			A	
🗀 Jobs	D Places	Name	Size	Modified 🔻 🔶		
	Search	Comorbidities.xls	11.3 kB	10/02/17		
	Recently Used	etl_mimic_20170215	557.0 kB	15/02/17		
	ү 🛅 data-integration	etl_mimic_20170313	1.3 MB	13/03/17	ot	
	🔊 revirier	etl_mimic_20170406	1.4 MB	06/04/17		
	🔯 Desktop	@ etl_mimic_20170428	1.5 MB	28/04/17		
	File System	@ etl_mimic_20170508	1.5 MB	lunes		
	🖾 data	etl_mimic_20170508_2	1.5 MB	lunes		
	Documents	© etl_mimic_20170509	1.6 MB	martes		
	🐻 Music					
	Dictures					
	🛅 Videos					
	💿 Downloads					
				*		
		Take your ETL. Analytics & Big Data to th	ne Next Level			
					•	
	4					

Contestamos no a la pregunta de si aplicamos un conjunto de reglas y aceptamos la instalación en el directorio raíz del repositorio:

8	Spoon - Welcome!	Xa-integr
File Edit View Action Tools	Repository import	×
View Design Explorer C C C C C C C C C C C C C C C C C C	Import job 17: J SELECT CASES UVI/USY11 USU/UJJJU SWC JAVEG TARSFORMATION #4 In the repository as "1_E LL_ASSIGN_CD9" 2017/05/11 00:07:04.373 is aved transformation #6 in the repository as "T_ETL_COMBINE_TMP_TABLES" 2017/05/11 00:07:04.573 is aved transformation #7 in the repository as "T_ETL_CD9_PROC" 2017/05/11 00:07:04.573 is aved transformation #7 in the repository as "T_ETL_CD9_PROC" 2017/05/11 00:07:04.573 is aved transformation #8 in the repository as "T_ETL_CU_DPROC" 2017/05/11 00:07:05.475 is aved transformation #10 in the repository as "T_ETL_CU_VENTS" 2017/05/11 00:07:05.475 is aved transformation #10 in the repository as "T_ETL_CU_VENTS" 2017/05/11 00:07:05.475 is aved transformation #11 in the repository as "T_ETL_CU_USTAYS" 2017/05/11 00:07:06.074 is aved transformation #13 in the repository as "T_ETL_CU_DAD SCORES" 2017/05/11 00:07:06.074 is aved transformation #16 in the repository as "T_ETL_CU_DAD SCORES" 2017/05/11 00:07:06.074 is aved transformation #16 in the repository as "T_ETL_CU_USTAYS" 2017/05/11 00:07:06.074 is aved transformation #16 in the repository as "T_ETL_CU_USTAYS" 2017/05/11 00:07:06.074 is aved transformation #16 in the repository as "T_ETL_CU_USTAYS" 2017/05/11 00:07:06.074 is aved transformation #16 in the repository as "T_ETL_CU_USTAYS" 2017/05/11 00:07:06.074 is aved transformation #16 in the repository as "T_EPL_CU_UCONSOLIDE_CLUCOSE" 2017/05/11 00:07:07.224 is aved transformation #20 in the repository as "T_EPL_CU_CU_UCOSE_TIMES" 2017/05/11 00:07:08.488 is aved transformation #21 in the repository as "T_EPL_CU_U_UCUSE_VARIAB" 2017/05/11 00:07:08.448 is aved is ansformation #23 in the repository as "T_EPL_CU_U_UCUSE_VARIAB" 2017/05/11 00:07:08.448 is aved is ansformation #23 in the repository as "T_EPL_CU_U_UCUSE_VARIAB" 2017/05/11 00:07:08.448 is aved job #1 in the repository as "T_EPL_VENTS" 2017/05/11 00:07:08.448 is aved job #1 in the repository as "T_EPL_UEU_NUTRITION" 2017/05/11 00:07:08.448 is aved job #25 in the repository as "T_EPL_VENTS"	
L		

Una vez terminado explorando el repositorio desde **Tools** -> **Repository** -> **Explore...**

8		Spoon - [mimiccat0	2] Welcome!			IX b-ij
File Edit View Action Too	ls Help	*	Repository ex	(plorer on [mimiccat02]		×
		Browse Connections	Hadoop Clusters Security	Slaves Partitions Clusters		$ \rightarrow$
🔄 View 🖉 Design	🔀 Welcome! 🛛	Folders: LaB=	Name	Туре	Date Modified	<u>^</u>
Explorer 🔍 🧹 🐁 🗄	🏽 🔹 👘 🕼 🛞 🍘	🗖 /	J_CREATE_TABLES_AD	JOB	8 May 2017 03:19:17 CEST	
Transformations			J_CREATE_TABLES_EVE	JOB	8 May 2017 03:28:48 CEST	
			J_CREATE_TABLES_GROUP	JOB	9 May 2017 02:11:01 CEST	
	Pentano L		J_CREATE_TABLES_INS	JOB	8 May 2017 03:28:57 CEST	
			b J_CREATE_TABLES_ME	JOB	8 May 2017 03:29:02 CEST	
	Wolcomo		J_CREATE_TABLES_SEL	JOB	9 May 2017 01:48:13 CEST	
	Welcome		Doc	JOB	6 Feb 2017 22:14:49 CET	
			퉎 J_ETL_ADM	JOB	9 May 2017 05:02:37 CEST	
			J_ETL_EVENTS	JOB	9 May 2017 03:29:44 CEST	
	Cot tho		🏠 J_ETL_INSULIN	JOB	9 May 2017 03:41:54 CEST	
	Gettie		J_ETL_MED_FEATURES	JOB	9 May 2017 03:44:44 CEST	
	From P		🂫 J_GLOBAL	JOB	9 May 2017 03:41:48 CEST	
			퉎 J_GLOBAL_ETL	JOB	9 May 2017 03:43:36 CEST	
	Let us help vo		🍃 J_GROUP_DATA	JOB	9 May 2017 02:46:34 CEST	
	an ETL. Big Da		🍃 J_GROUP_DATA_GLUCC	JOB	9 May 2017 01:55:35 CEST	
			🏠 J_POST_ETL	JOB	9 May 2017 02:57:01 CEST	
			J_SELECT_CASES	JOB	9 May 2017 02:53:35 CEST	
	Tutoriais & vide		№ т_doc	TRANSFORMATION	8 May 2017 01:29:48 CEST	
			T_ETL_ADT_ALLERGIES	TRANSFORMATION	8 May 2017 03:54:17 CEST	
			T_ETL_ADT_SEPSIS	TRANSFORMATION	9 May 2017 03:45:04 CEST	
			T_ETL_ASSIGN_COMOR	TRANSFORMATION	9 May 2017 04:49:25 CEST	
	-		T_ETL_ASSIGN_ICD9	TRANSFORMATION	9 May 2017 04:54:34 CEST	•
					C	lose

comprobamos que se han cargado correctamente las transformaciones y trabajos.

Cargando cualquier trabajo, por ejemplo, el inicial J_GLOBAL, podemos ver en el panel de la izquierda las conexiones de datos que es necesario configurar con los datos específicos de cada entorno: **mimic_in** que se corresponde con la BBDD MIMIC-III, **mimic_out** que es la BBDD intermedia *mimicsel* que hemos creado anteriormente y **temp_db** BBDD intermedia en memoria que no es necesario reconfigurar.

A5-2. INSTALACIÓN MIMIC-III

Para la instalación de la BBDD MIMIC-III existen tres alternativas, documentadas en el sitio web de MIMIC (<u>https://mimic.physionet.org/</u>) y en MIMIC-CODE (<u>https://github.com/MIT-LCP/mimic-code</u>)

- Instalar PostgreSQL, obtener los ficheros CSV con el contenido de las tablas, crear una BBDD vacia y seguir los pasos detallados en <u>https://mimic.physionet.org/tutorials/install-mimic-locally-ubuntu/</u> para cargar los ficheros CSV en una nueva base de datos.
- Instalar PostgreSQL (u otro gestor de BBDD soportado), obtener los ficheros CSV con el contenido de las tablas, clonar el repositorio mimiccode y seguir las instrucciones de <u>https://github.com/MIT-LCP/mimiccode/tree/master/buildmimic/postgres</u> para ejecutar los scripts de instalación.
- Instalar PostgreSQL y MIMIC-III en una máquina virtual mediante Vagrant siguiendo las instrucciones del repositorio <u>https://github.com/nsh87/mimic-iii-vm</u>

Las dos últimas opciones automatizan todo el proceso, la carga de los ficheros CSV en la BBDD PostgreSQL tarda unas 7 u 8 horas dependiendo de las características del equipo y la configuración del gestor de BBDD.

Las vistas materializadas se pueden instalar ejecutándose desde cualquier entorno de consultas SQL conectado a la BBDD MIMIC-III o siguiendo las instrucciones del repositorio <u>https://github.com/MIT-LCP/mimic-</u> code/blob/master/Makefile.md