Data Governance- Talend Data Catalog

BIG DATA – BUSINESS INTELLIGENCE – MACHINE LEARNING

CONTENIDO

1. GOBIERNO DEL DATO

1.1 INTRODUCCIÓN

En el presente documento se pretende dar una pincelada al gobierno del dato, destacando los conceptos principales y resumiendo ciertos aspectos a tener en cuenta para llevar a cabo un proyecto de este tipo. Se propone el ecosistema de herramientas de Talend como software de apoyo para el desarrollo.

Dentro de un concepto global Data Management, el concepto de Data Governance se encarga de la dirección y supervisión de los datos, identificando además las siguientes áreas que están interrelacionadas:

- Arquitectura de datos: Define el plan de gestión de los activos de datos alineándose con la estrategia organizativa para establecer los requisitos de datos estratégicos y los diseños para cumplir con estos requisitos.

- Modelización y diseño de datos: Es el proceso de descubrir, analizar, representar y comunicar los requisitos de datos en una forma precisa llamada modelo de datos.

- Almacenamiento y operaciones de datos: Incluye el diseño, la implementación y el mantenimiento de los datos almacenados para maximizar su valor. Las operaciones se realizan a lo largo del ciclo de vida de los datos, desde la planificación hasta la eliminación de los mismos.

- Seguridad de los datos: La seguridad de los datos garantiza que la privacidad y la confidencialidad de los datos se mantengan, que no se violen los datos y que se acceda a ellos de manera adecuada.

- Integración e interoperabilidad de datos: Incluye procesos relacionados con el movimiento y consolidación de datos dentro y entre almacenes de datos, aplicaciones y organizaciones.

- Gestión de documentos y contenidos: Incluye las actividades de planificación, implementación y control utilizadas para gestionar el ciclo de vida de los datos y la información que se encuentran en una serie de medios, especialmente los documentos necesarios para apoyar los requisitos de cumplimiento legal y reglamentario.

- Datos de referencia y maestros: Incluye la conciliación y el mantenimiento continuos de datos compartidos fundamentales para permitir el uso coherente en todos los sistemas de la versión más exacta, oportuna y pertinente de la verdad sobre las entidades comerciales esenciales.

- Data warehousing y BI: Incluye los procesos de planificación, ejecución y control para gestionar los datos de apoyo a la toma de decisiones y permitir a los analistas de datos obtener valor de los datos mediante el análisis y la presentación de informes.

- Metadatos: Incluye las actividades de planificación, ejecución y control para permitir el acceso a metadatos integrados y de alta calidad que incluyen definiciones, modelos, flujos de datos y otra información crítica para comprender los datos y el sistema a través del cual se crean, se mantienen y se accede a ellos.

- Calidad de los datos: Incluye la planificación y aplicación de técnicas de gestión de la calidad para medir, evaluar y mejorar la idoneidad de los datos para su uso dentro de una organización.

Ilustración 1. DAMA-DMBOK2 Framework

Existen diversos frameworks que nos pueden ayudar a la hora de abordar un proyecto de este tipo estableciendo criterios, equipo de trabajo, actividades a desarrollar, entregables y buenas prácticas para la correcta gestión del dato como son DAMA, DGI o DCAM

1.2. EQUIPO DE TRABAJO, ACTIVIDADES Y ENTREGABLES

En este punto mencionaremos tres aspectos importantes a destacar.

- Equipo de trabajo: En lugar de centrarnos en "títulos" de puestos concretos, vamos a destacar las habilidades que debemos reunir para cubrir todas las áreas a desarrollar:
    - Procesos de negocio: Comprender los requisitos y determinar el impacto en los datos
    - Modelado y arquitectura de datos: Establecer un roadmap y arquitectura
    - Habilidades técnicas de arquitectura: Diseño de dw, construcción y gestión
    - Manipulación de datos: Resolución de problemas estructurales, relación entre datos y datasources
    - Analítica de datos: Interpretación de datos, representación gráfica
    - Habilidades con el idioma: Definir glosarios comprensibles
    - Habilidades estratégicas de negocio: Comprender las necesidades de la organización
- Actividades:
    - Evaluación de la situación actual, identificación de necesidades e identificación de casos de uso
    - Definición de la estrategia y el framework operativo centrado en casos de uso
    - Gestión del cambio, creación del glosario de negocio (business glosary) y coordinación con otras áreas
    - Hacer operativo el modelo para toda la organización
- Entregables
    - Plan estratégico y roadmap
    - Análisis de procesos
    - Framework operativo y roles de trabajo
    - Glosario de negocio

- Cuadro de mando resumen
- Plan de gestión del cambio, comunicación y formación
- Buenas prácticas, guías

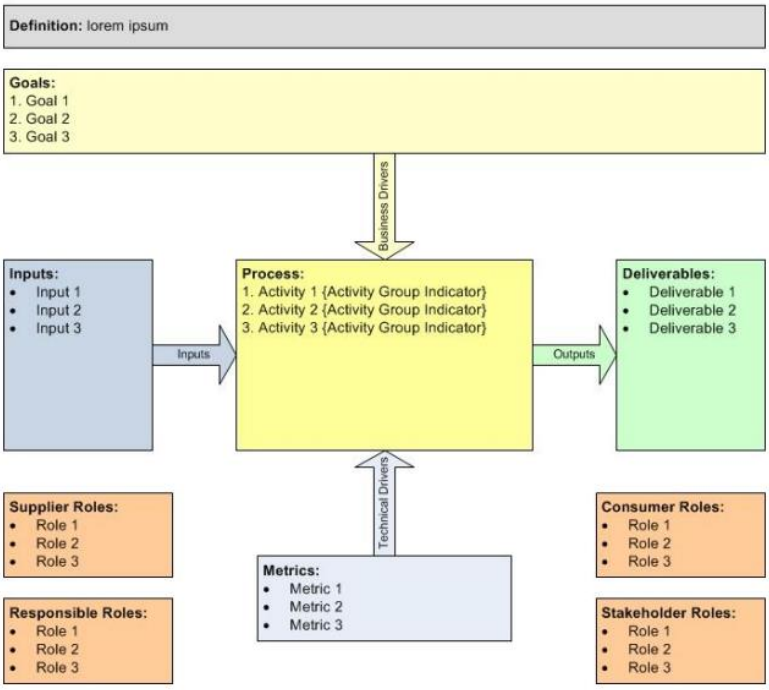


Ilustración 2. Template para documentación proceso de negocio

2. TALEND DATA CATALOG

Talend ofrece un conjunto de herramientas que nos permite cubrir todo el ciclo de vida de los datos. En el siguiente enlace un vídeo que muestra toda su capacidad. La documentación completa se puede consultar en este enlace.

En este documento nos centraremos en la herramienta Talend Data Catalog que nos permite definir un catálogo central y gobernado de datos de confianza.

Dicho catálogo puede ser compartido y elaborado de manera colaborativa fácilmente. Puede descubrir, perfilar, organizar y documentar automáticamente sus metadatos y hace que sea fácil de buscar.

Tal y como se menciona en el punto anterior, el data governance es solo una pieza dentro del data management y por tanto, no hay que restar importancia a otros procesos como:

- Data Quality: limpieza, integración, profiling
- Definir los BPM de nuestro negocio
- Gestión de los workflows y responsabilidades entre los roles de la organización

Para ello Talend proporciona también otras herramientas como Talend Data Preparation y Talend Data Stewardship. Más info en el siguiente enlace

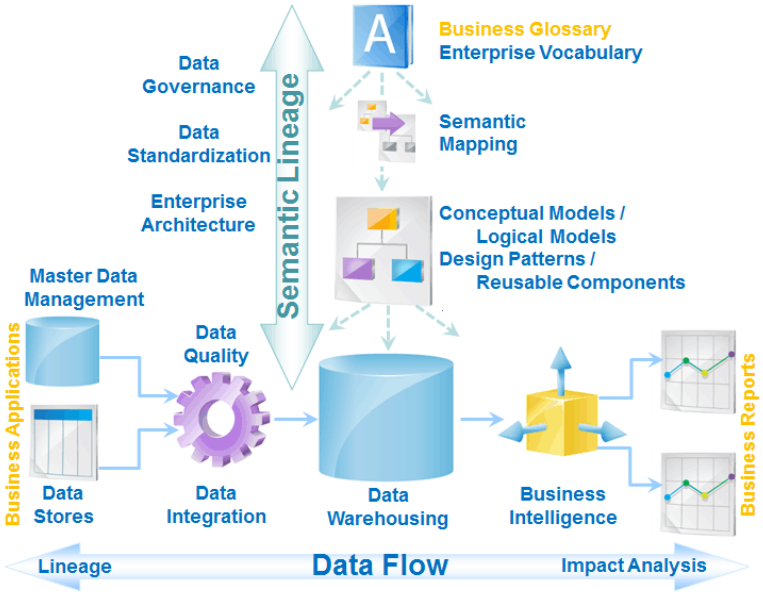


Ilustración 3. Metadata Management Arquitecture

2.3 PRINCIPALES CARACTERÍSTICAS

- Análisis automático de datos. Muestra patrones y estadísticas básicas de los datos.
- Estudio del linaje y análisis del impacto.
- Recolección automática y programable de metadatos desde más de cien aplicaciones distintas. Consultar aquí
- Validación de datos mediante diccionarios semánticos. También permite el enmascaramiento de los datos.
- Gestión de la seguridad de los archivos permitiendo acceso en función de roles y permisos predefinidos a usuarios y grupos.
- *Social catalog* es un conjunto de características que permiten el trabajo en equipo facilitando la comunicación mediante comentarios, avisos…
- Gestión de los datos mediante glosarios que pueden ser trabajados con workflow.

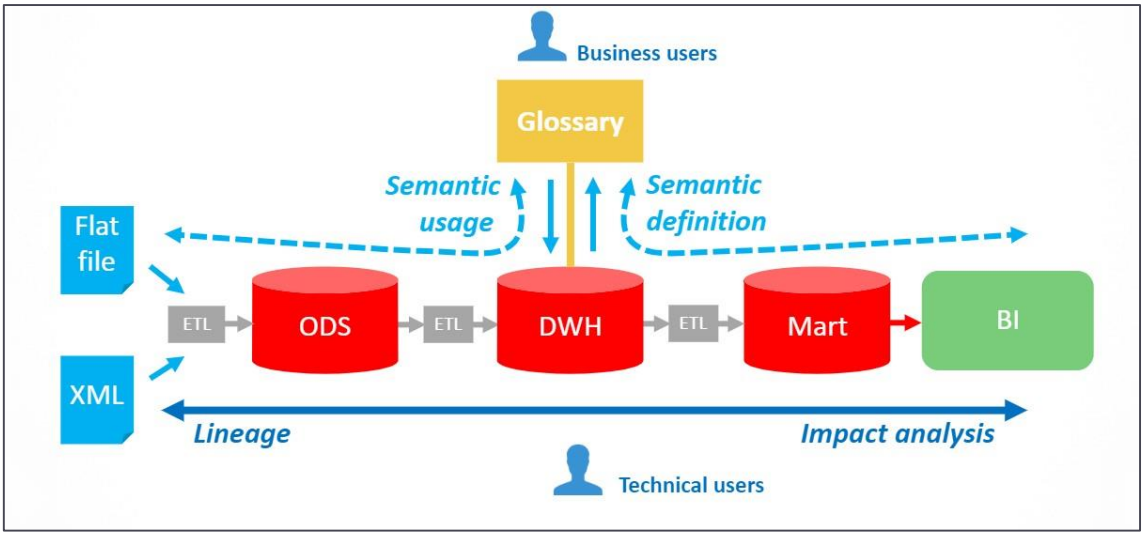
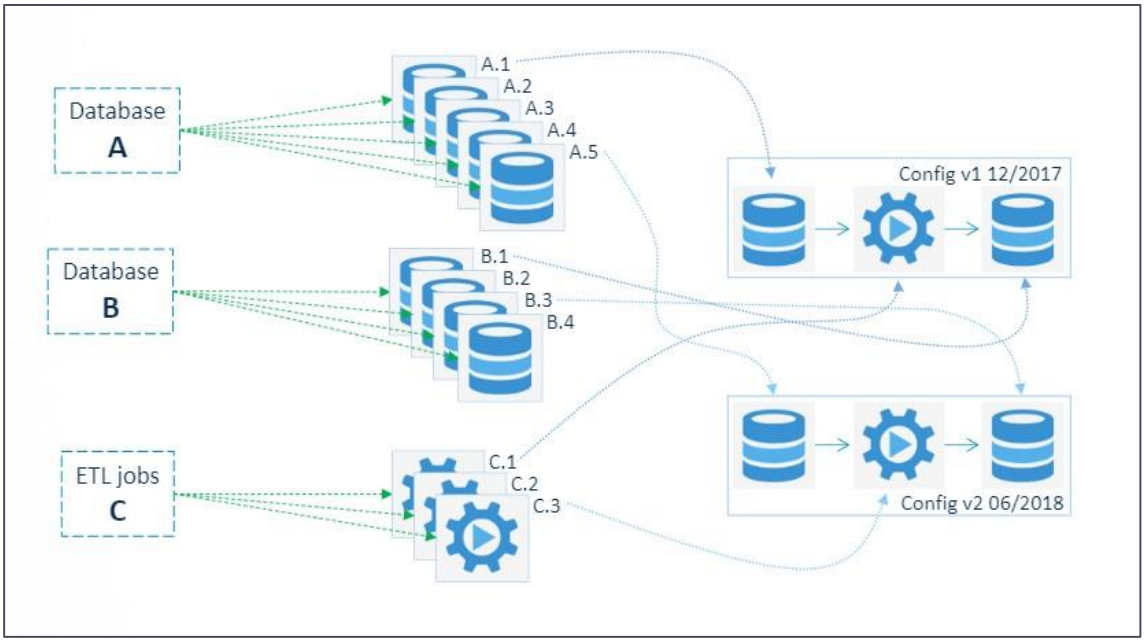


Ilustración 4. Ejemplo de Data Workflow con Glosario

- Internamente *Talend Data Catalog* trabaja con versiones de metadatos, por defecto están ocultas, pero pueden hacerse visibles desde la interfaz de administración.



- Ilustración 5. Ejemplo de Control de Versiones

  - La interfaz *Metadata Explorer* muestra las distintas configuraciones creadas en la interfaz *Metadata Manager*.

Se pueden modificar los menús y el aspecto alterando el archivo *MetadataExplorar.xml* o desde el propio *Metadata Explorer:*

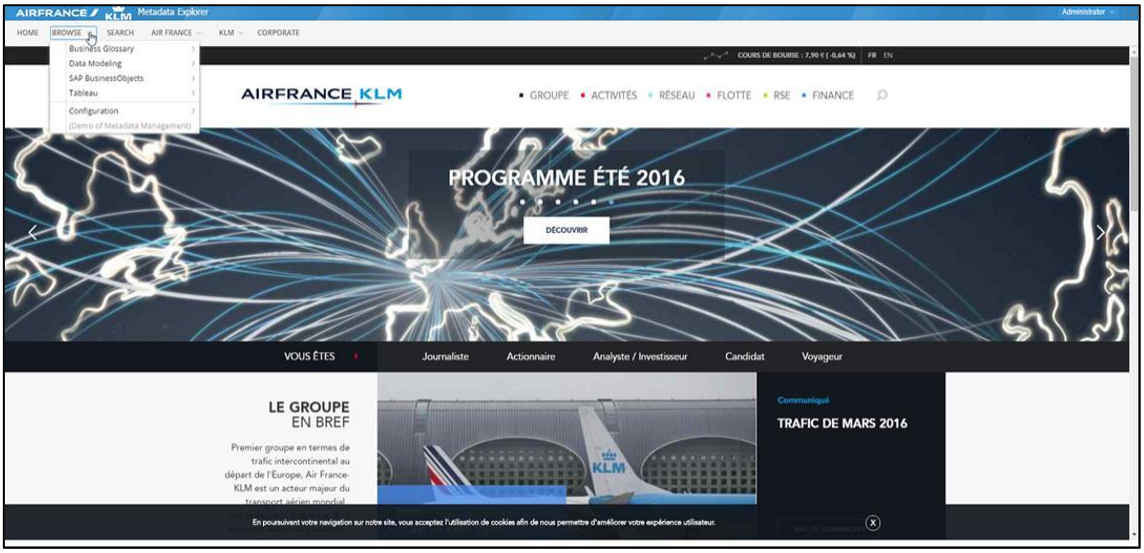


Ilustración 6. Ejemplo de interfaz Metadata Explorer personalizada

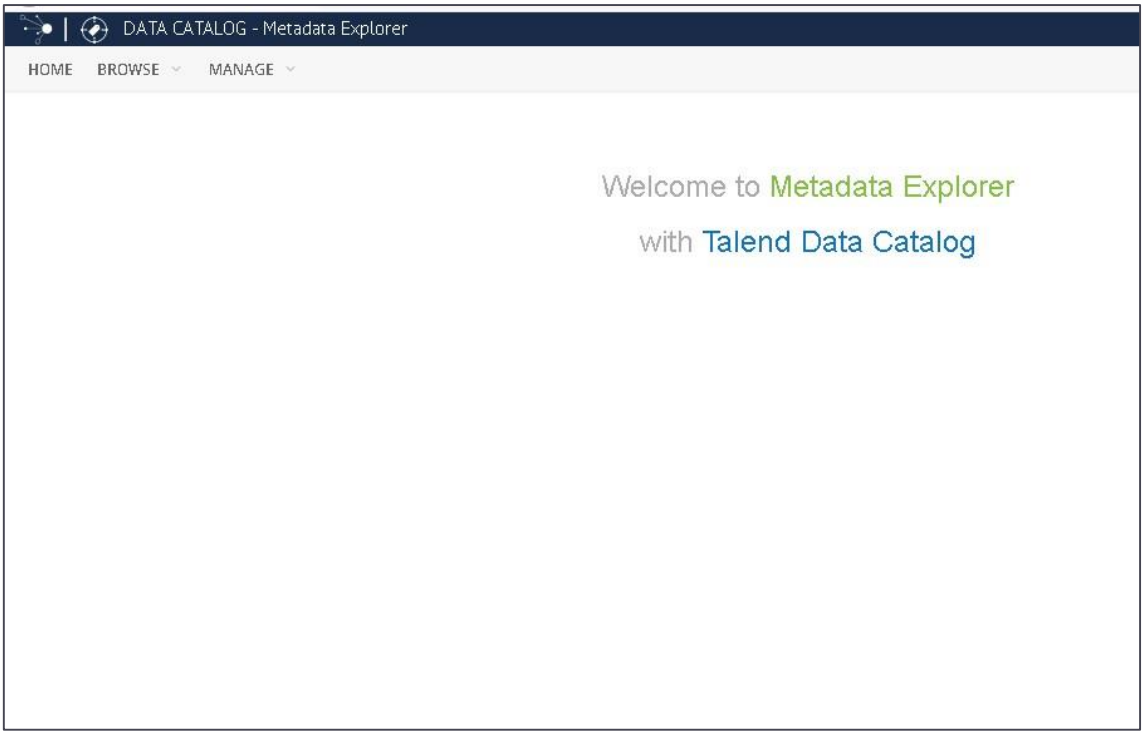


Ilustración 7. Interfaz Metadata Explorar por defecto

2.1 ARQUITECTURA Y DESCRIPCIÓN DEL SERVICIO

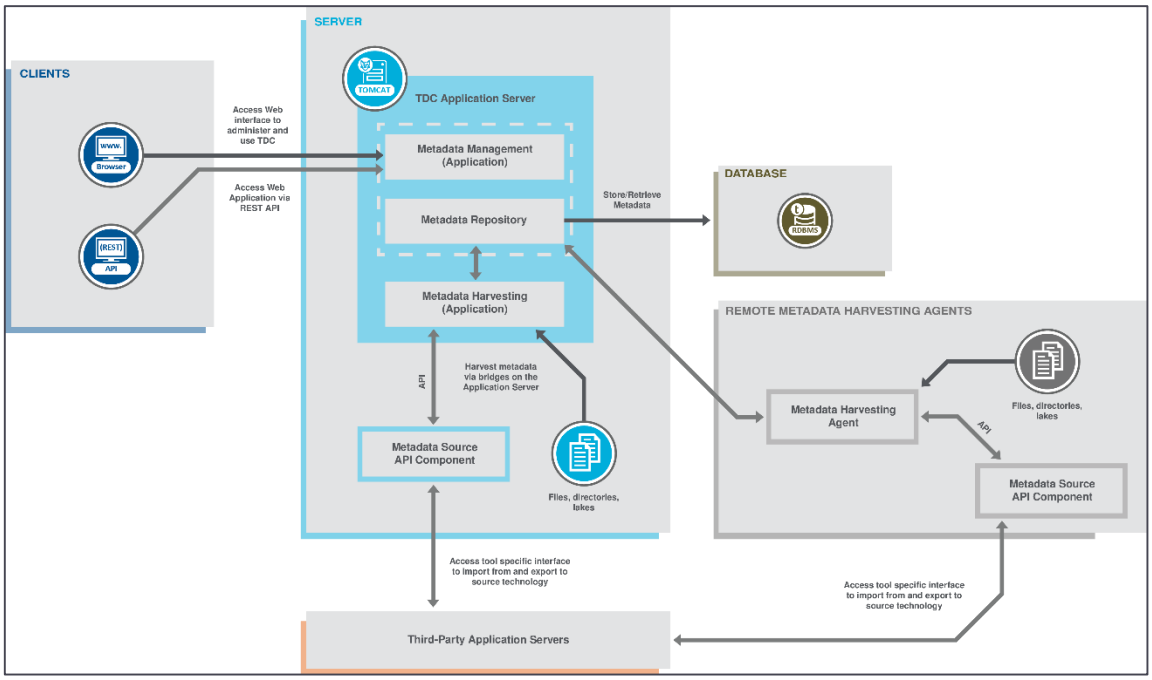


Ilustración 3. Ejemplo de arquitectura Talend Data Catalog

Descripción de los bloques:

- **Clientes**: se conectan con un navegador web y una API REST. Desde el navegador Web o la API REST, se accede a la aplicación *Talend Data Catalog*. Aquí es donde se crean los modelos, se importan los metadatos, se rastrea el linaje, se gestiona el repositorio de metadatos, se diseña la arquitectura de la empresa y se gestionan las tareas administrativas.

- **Servidor**: incluye la aplicación Web *Talend Data Catalog*. Utiliza un servidor Apache Tomcat integrado y se ejecuta como una aplicación web estándar. Por defecto, este servidor es accesible en el puerto 11480.

La aplicación Web de *Talend Data Catalog* incluye:

- **Metadata Management** para la gestión de los metadatos, el gobierno de los datos y la catalogación de los datos, el Depósito de Metadatos para el almacenamiento de metadatos, la aplicación de recolección de metadatos para la importación y exportación de metadatos.

- **Base de datos**: contiene la base de datos utilizada para almacenar y recuperar metadatos y ejecutar cálculos sobre ellos. Las bases de datos admitidas son Oracle, PostgreSQL y MS SQL Server.

- **Remote Metadata Harvesting Agents**: Uno o más servidores de recolección de metadatos instalados en una máquina remota. *Talend Data Catalog* utiliza esta máquina como agente para importar o exportar a la tecnología de la fuente de metadatos.

2.2 INTERFACES

Talend Data Catalog cuenta con dos interfaces a las cuales se accede por un punto de entrada común, en función de los permisos de los que se disponga se mostrará la interfaz de usuario estándar o la interfaz de administración.

- **Metadata explorer:** Permite explorar los metadatos, así como realizar modificaciones sobre estos en función de los permisos de los que se dispongan. Si se accede desde una cuenta de administrador aparece un nuevo menú desde el cuál se realiza una gestión más amplia de los metadatos, los permisos de usuario, los servidores, y lo más importante, se abre el acceso a la interfaz de administración.

- **Metadata manager:** En esta interfaz se realiza la gestión completa tanto de los metadatos como de las vistas de los mismos disponibles desde la interfaz de exploración. Permite la realización de copias de seguridad y versionado de los metadatos, así como gestionar la seguridad de los mismos.

2.4 RECOLECCIÓN DE DATOS EN TALEND DATA CATALOG

Talend Data Catalog permite obtener metadatos desde distintas plataformas mediante el uso de puentes propios, a los que se suman el código de terceros para ampliar esta funcionalidad.

Para importar metadatos en *Talend Data Catalog* es necesario hacerlo desde una cuenta del tipo administrador. Desde el menú *Manage* accedemos a *Contents*:

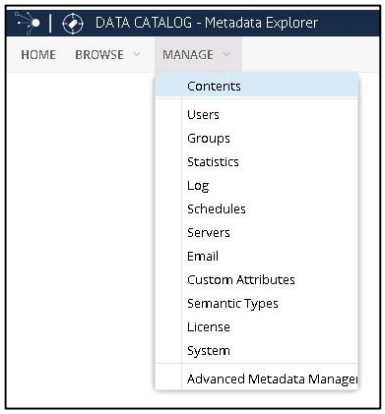


Ilustración 8. Administración de Contenidos

Una vez accedido a la gestión de contenidos presionamos el símbolo + y seleccionamos *Model:*

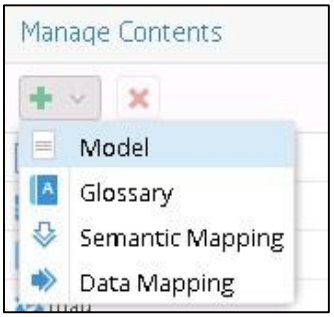


Ilustración 9. Inserción de nuevo Modelo

Una vez hemos accedido a M*odel* nos aparecerá una pestaña en la que seleccionaremos un puente en función de la fuente de la que queramos:

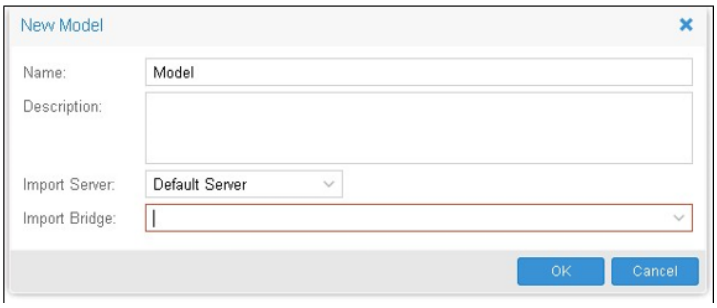


Ilustración 10. Selección del Puente

2.5.1 EXTRACCIÓN DE METADATOS DE UN ARCHIVO

Habiendo seguidos los pasos anteriormente mencionados, seleccionamos el puente *"File System (CSV, Excel, XML, JSON, Avro, Parquet, ORC, COBOL, Copybook)"*, introducimos el nombre y la descripción deseada y pulsamos *OK*:

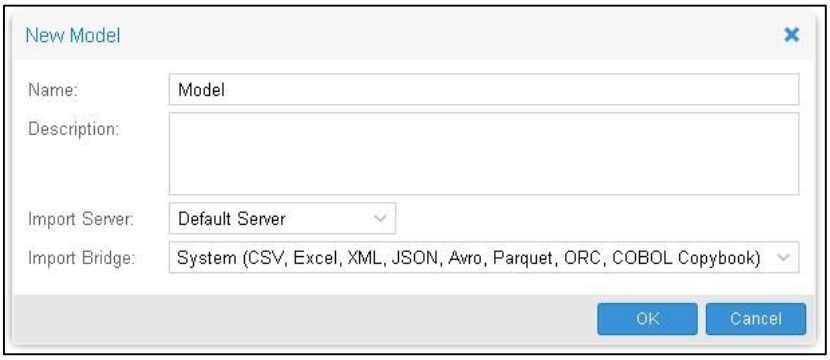


Ilustración 11. Selección de puente *File System*

Aparecerá un nuevo elemento y su correspondiente pantalla de configuración:

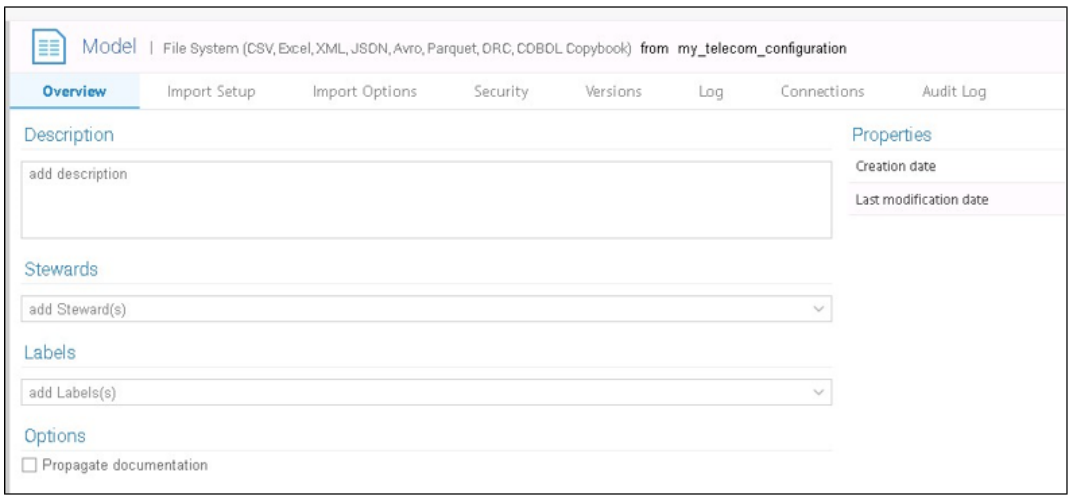


Ilustración 12. Vista Previa nuevo modelo

Desde la pestaña *Import Setup* seleccionamos el directorio y el o los ficheros:

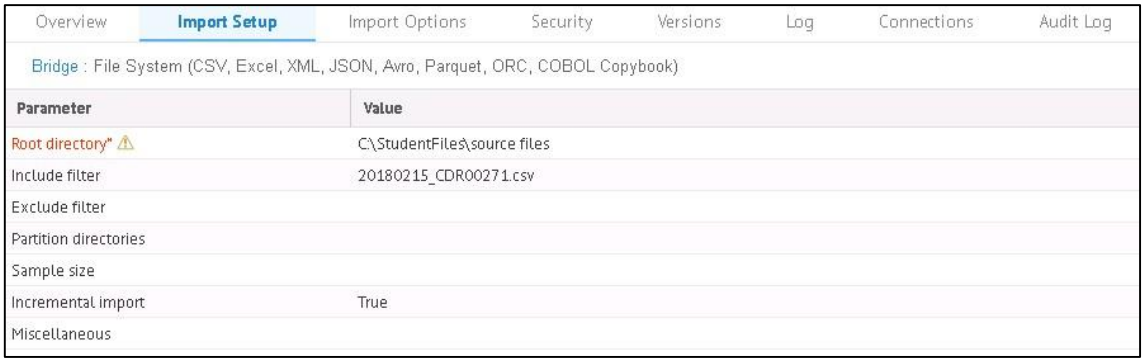


Ilustración 13. Preparación de importación de un Modelo de Fichero

Desde la pestaña *Import Options* podemos definir el número de filas analizadas para tomar los metadatos y el número de filas que tomará como muestra. Tras editar ambos valores pulsamos *Save* en la esquina superior derecha para confirmar los cambios:

Ilustración 14. Configuración de importación de un Modelo

Tras haber configurado el fichero, pulsamos *Import* para extraer los metadatos del fichero tras lo que aparecerá una ventana que mostrará el estado de la importación:

Ilustración 15. Ejecución de importación de Modelo

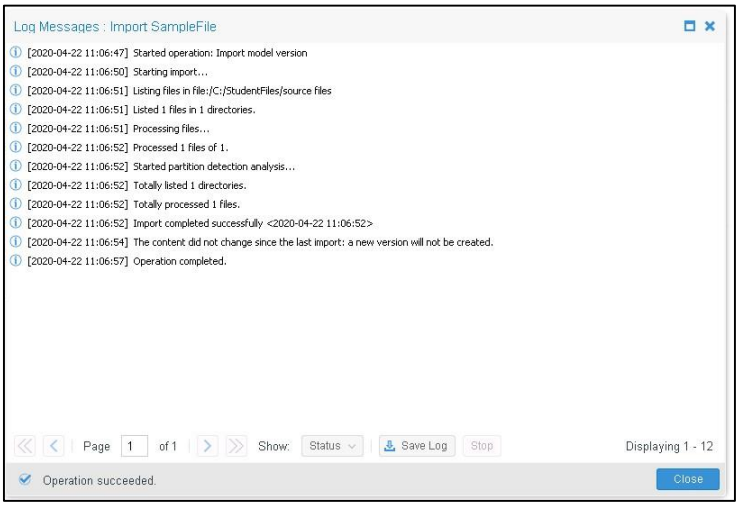


Ilustración 16. Log de importación de Modelo

Una vez finalizada la operación, podemos recurrir al menú *Browse* para asegurarnos de que los metadatos han sido correctamente introducidos:

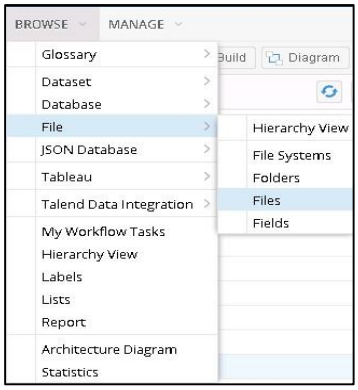


Ilustración 17. Menú de búsqueda de ficheros

Seleccionando el archivo que hemos importado podremos acceder a la visualización de los datos:

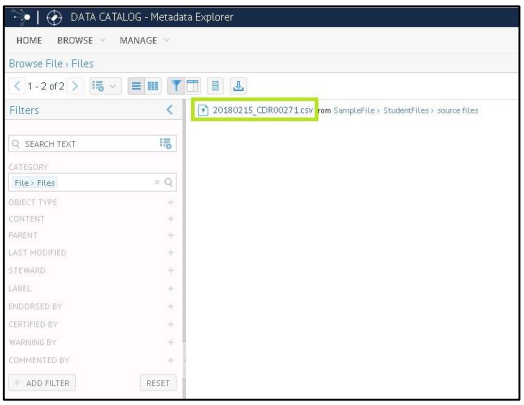


Ilustración 18. Selección de archivo importado

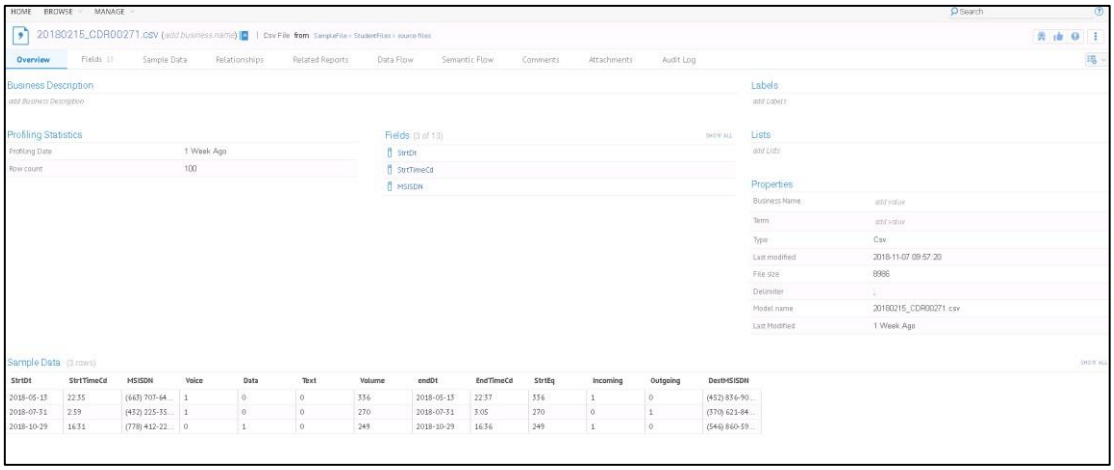


Ilustración 19. Vista previa de archivo

En este punto se muestra una vista general de los datos pero se puede realizar un análisis más detallado. En la pestaña *Fields* se muestran todos los campos, y si pulsamos encima de uno de ellos, se muestra un pequeño análisis de los datos del mismo:

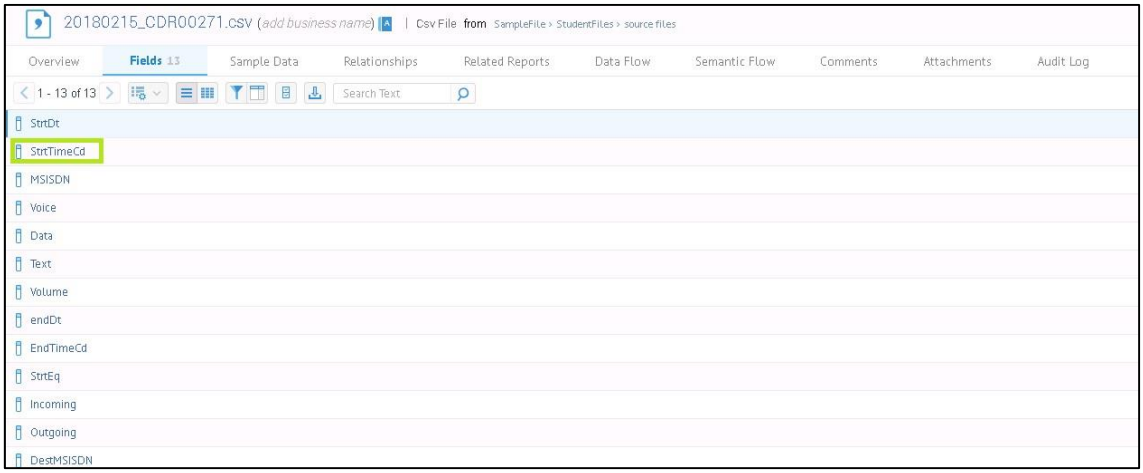


Ilustración 20. Campos del Archivo

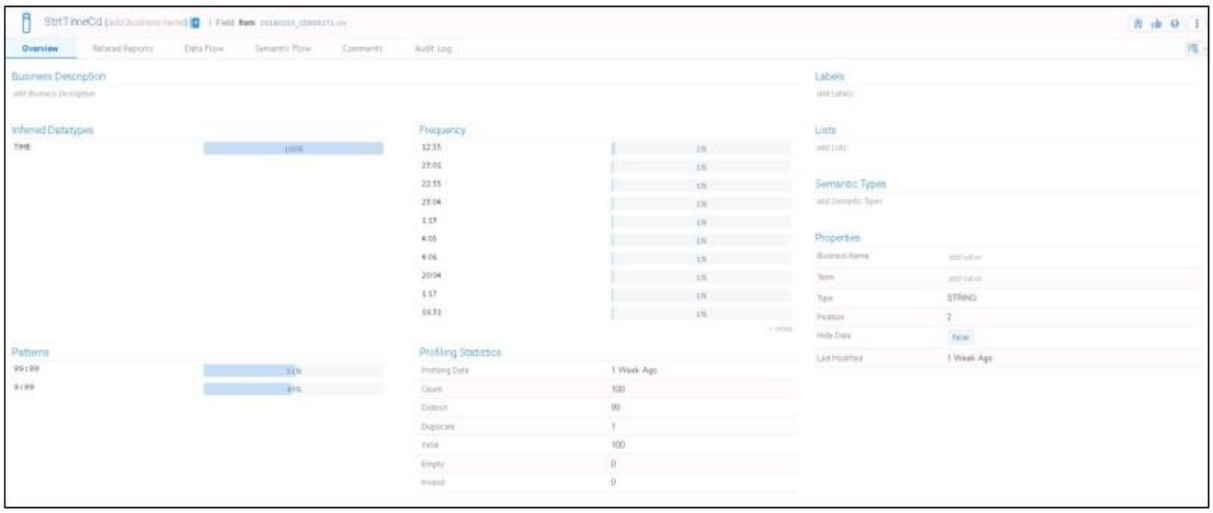
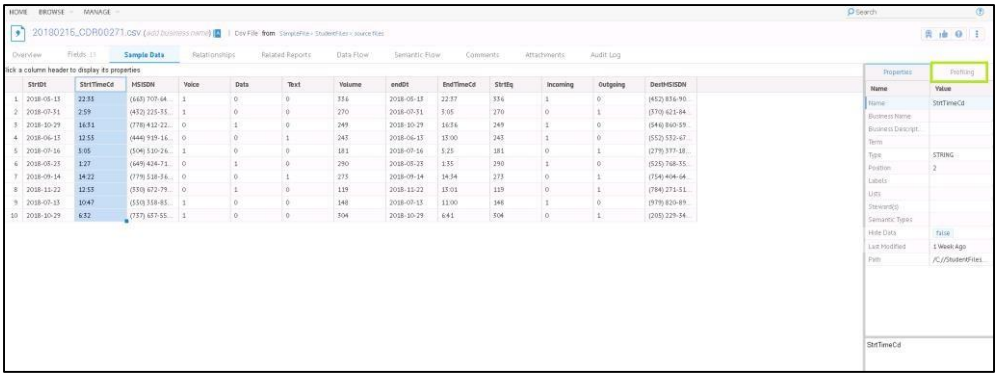
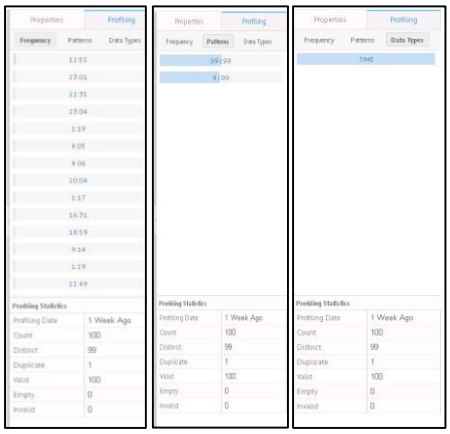


Ilustración 21. Vista previa de campo

En la pestaña *Sample Data* encontramos la muestra de datos que hemos establecido en la importación, desde donde también se puede ver un pequeño análisis de los datos. Por defecto se muestra la pestaña *Properties,* pero podemos cambiar a la pestaña *Profiling* donde se muestran más datos:





Ilustraciones 22-25. Perfil de datos de campo

2.5.2 EXTRACCIÓN DE METADATOS DE UNA BASE DE DATOS

TDC dispone de gran variedad de puentes para conectarse a bases de datos, en este ejemplo utilizaremos el puente *"Oracle MySQL Database (via JDBC)":*

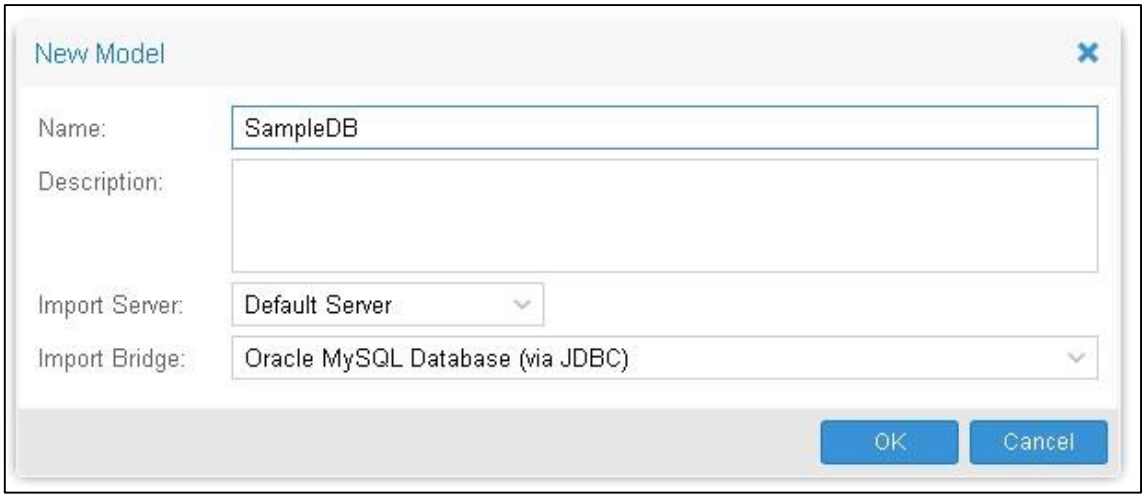


Ilustración 26. Selección de puente *Oracle MySQL Database*

Análogamente al ejemplo anterior, al crear un nuevo modelo se nos muestra la vista previa del componente. Desde el menú accedemos a *Import Setput* y añadimos los detalles de conexión a la base de datos:

Ilustración 27. Preparación de importación de un Modelo de Base de Datos

Una vez establecida la conexión, podemos seleccionar el o los esquemas de los que deseamos extraer los datos:

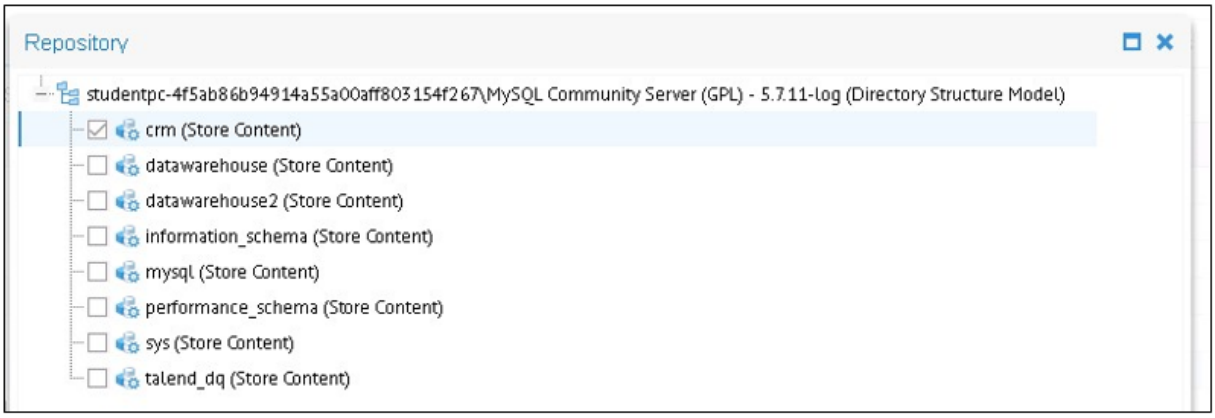


Ilustración 28. Selección de esquema

Tras seleccionar el esquema, probamos que funcione correctamente la conexión y guardamos:

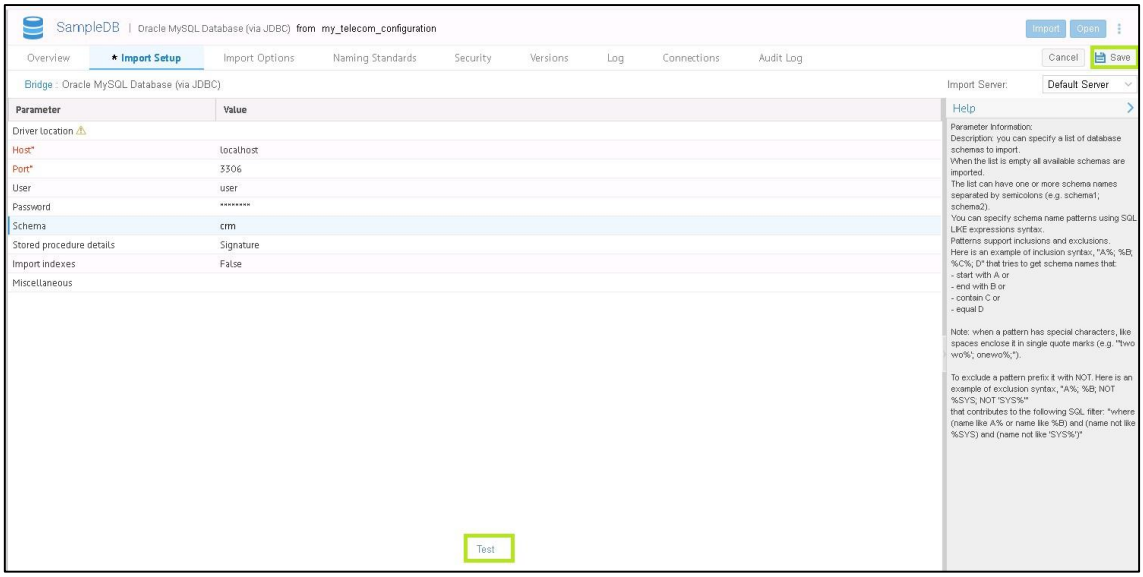


Ilustración 29. Prueba de conexión a Base de Datos

Cambiamos a la pestaña *Import Options* para configurar el número de filas que toma para establecer los metadatos y el número de filas que compondrán la muestra. Tras configurarlo guardamos la configuración e importamos los metadatos:

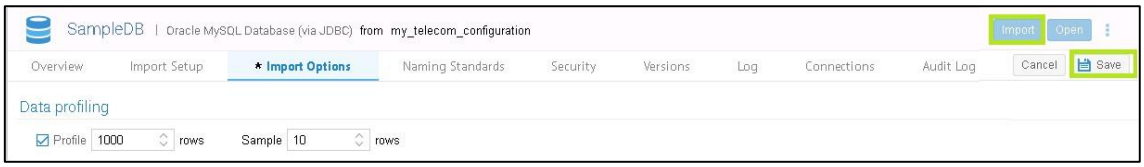


Ilustración 30. Configuración de importación de un Modelo

Con la importación aparecerá un log que mostrará el desarrollo de la operación:

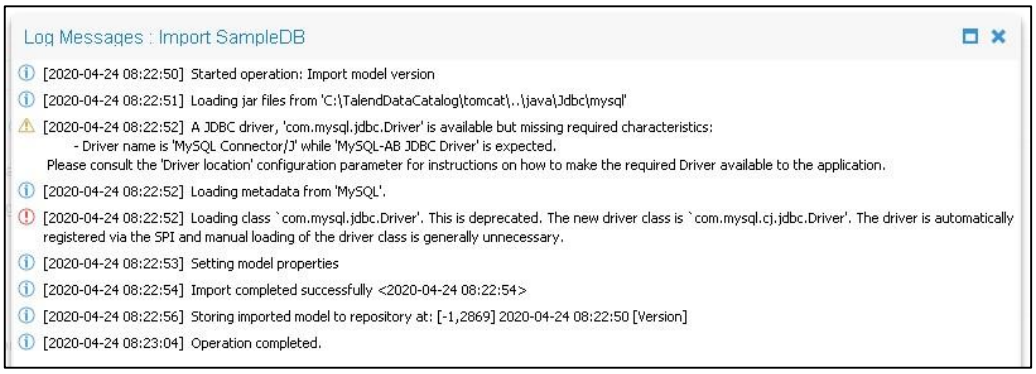


Ilustración 31. Log de importación de Modelo

Desde el menú *Brawse>Database>HierarchyView* podemos visualizar el contenido que acabamos de importar:

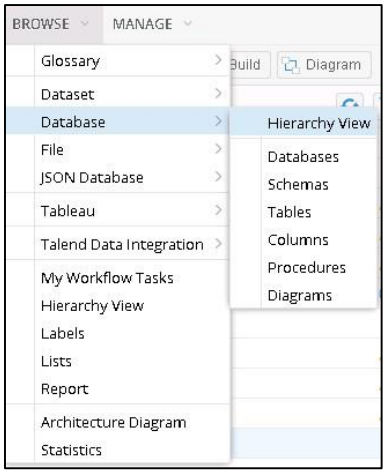


Ilustración 32. Menú de búsqueda de bases de datos

Una vez hecho esto, nos parecerá el esquema del modelo SampleDB recién importado, mediante el símbolo + podemos expandir el esquema tal y como se muestra en la siguiente ilustración:

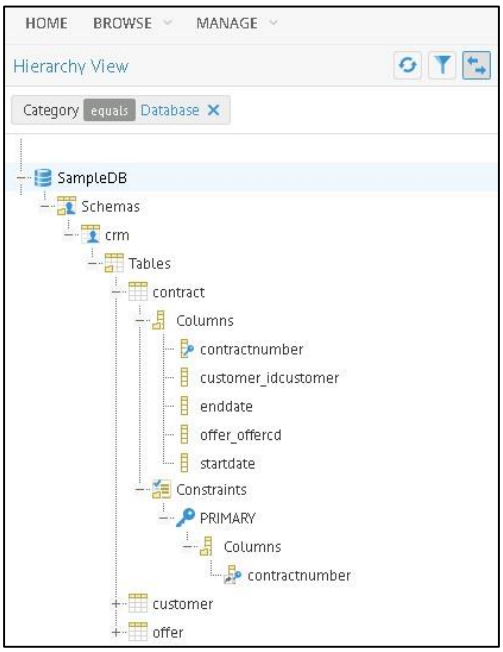


Ilustración 33. Esquema expandido de base de datos

Seleccionando cualquier campo podemos acceder a una vista que muestra los metadatos recopilados de dicho campo:

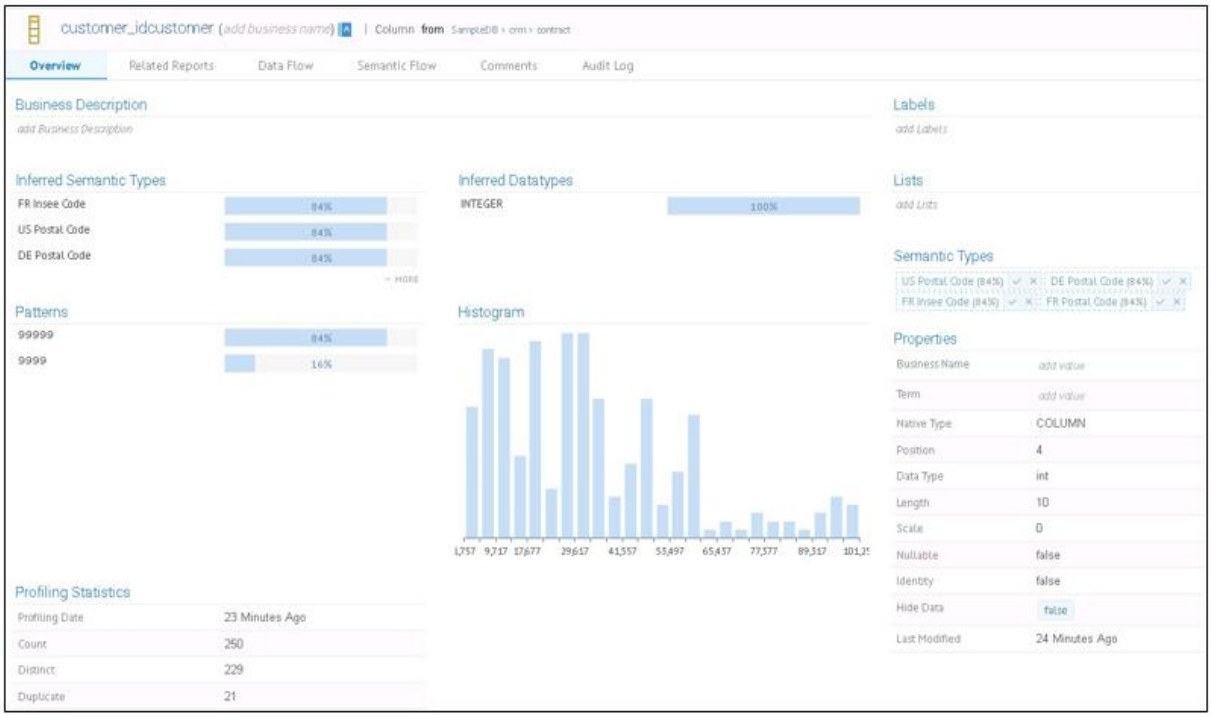


Ilustración 34. Vista previa de campo

2.5.3 EXTRACCIÓN DE METADATOS DE DATA INTEGRATION JOB

TDC dispone de puentes para conectar con otras herramientas Talend:

Ilustración 35. Puentes disponibles de conexión a herramientas Talend

Seleccionamos el puente que nos interese, en este ejemplo será *Talend Data Integration*:

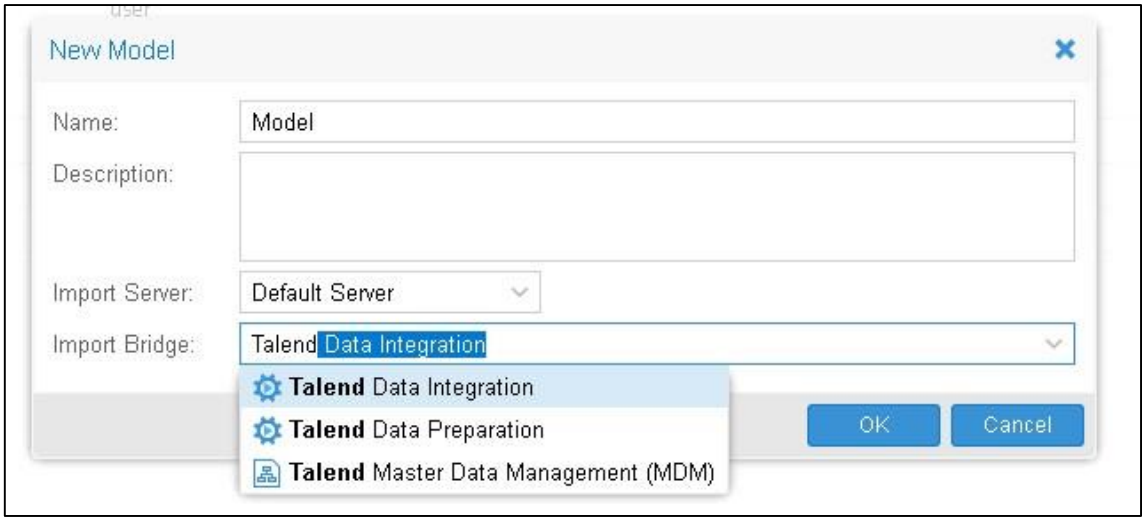


Ilustración 36. Creación de modelo utilizando puentes a herramientas Talend

Desde la pestaña *Import Setup* seleccionamos el directorio en el que se almacena el proyecto desde el cual queremos realizar la importación y seleccionamos el o los *ítems* que queremos importar desde el repositorio:

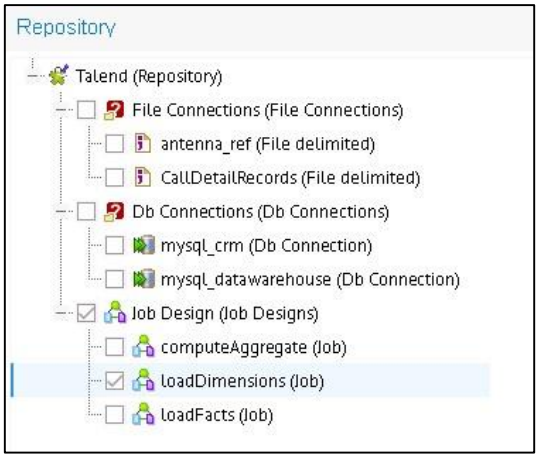


Ilustración 37. Selección de Trabajo *loadDimensions*

Una vez seleccionados, comprobamos la conexión, guardamos los cambios realizados e iniciamos el proceso de importación:

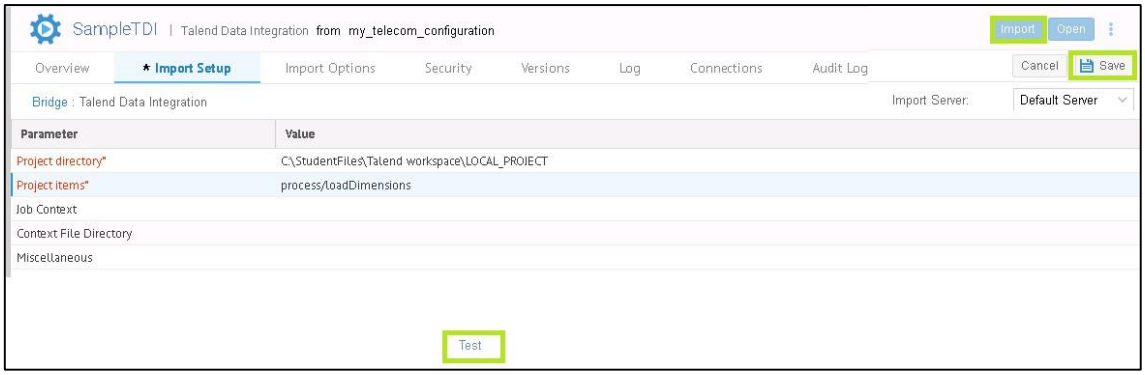


Ilustración 38. Preparación de importación de un Modelo TDI

Al realizar la importación aparece un log de la misma informando del progreso del proceso:

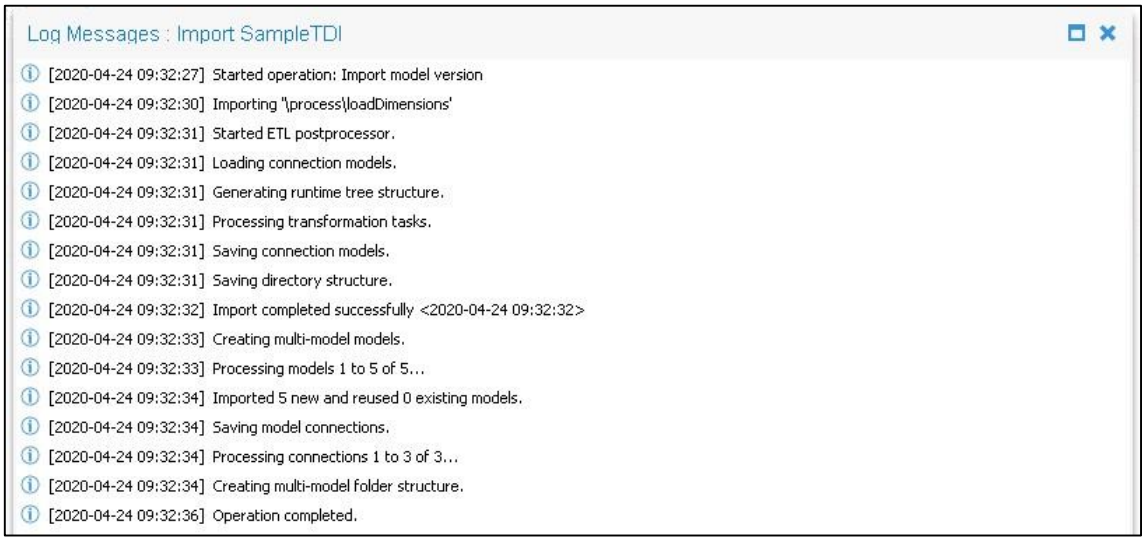


Ilustración 39. Log de importación de Modelo

Desde el menú *Browse>Talend Data Integration>HierarchyView* accedemos a la lista de los modelos TDI importados donde pulsando el icono + que aparece a la izquierda de cada componente podemos expandir el esquema:

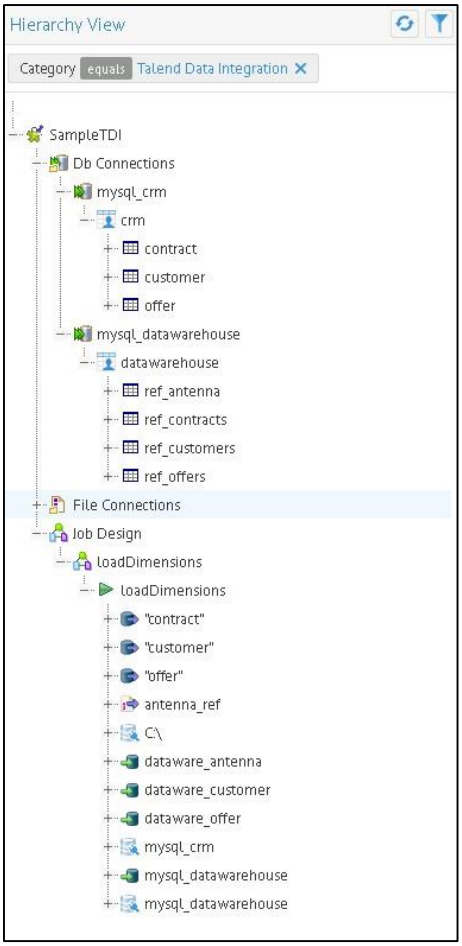


Ilustración 40. Esquema Modelo TDI importado

En cada componente encontramos distinta información referente al mismo. Accediendo al componente loadDimensions, que representa el trabajo cargado, podemos visualizar el *DataFlow* del mismo:

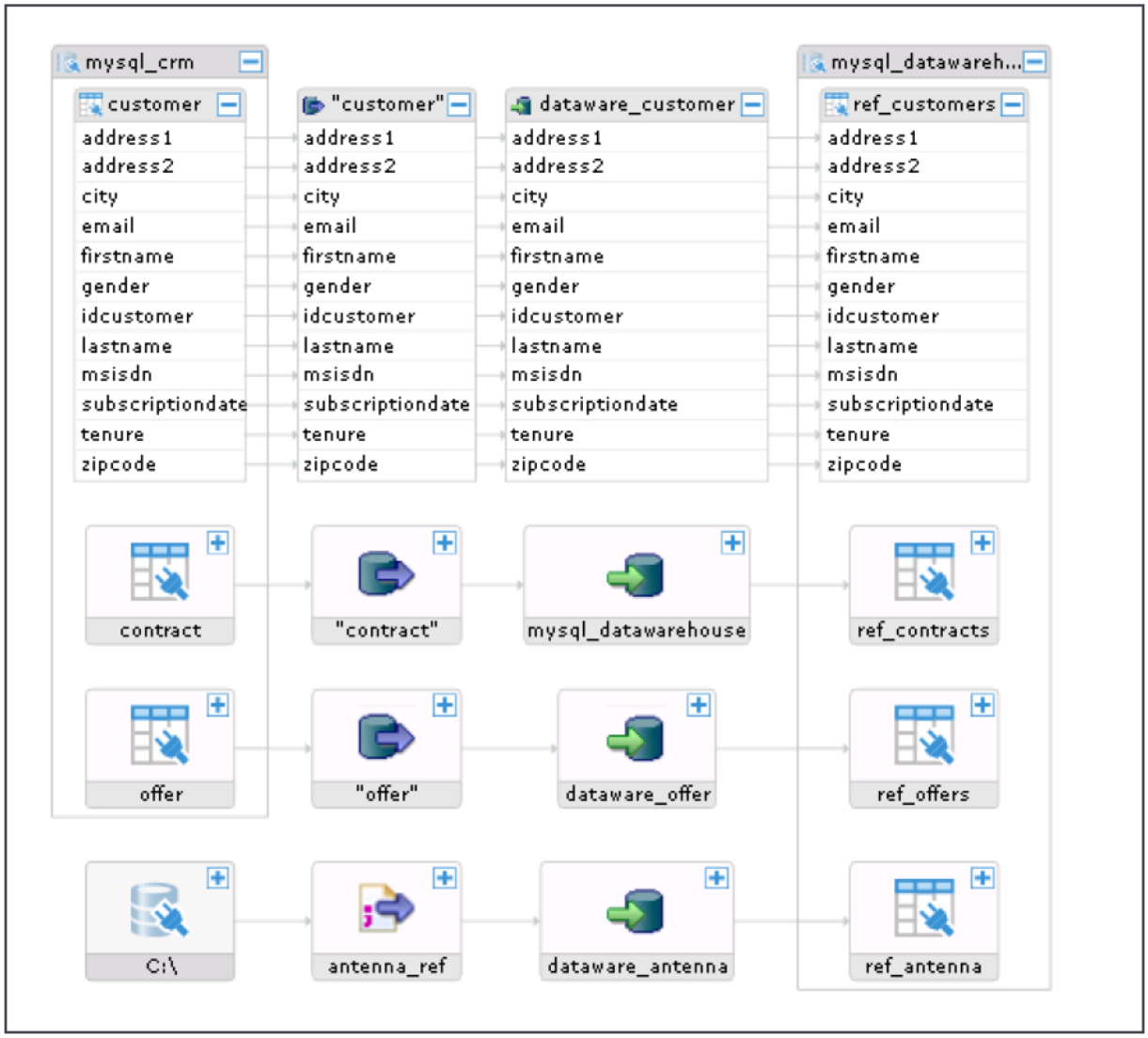


Ilustración 41. DataFlow del trabajo loadDimensions

TDC ha creado un vínculo con este archivo por lo que, si sufriera modificaciones, al recolectar nuevamente los datos (se puede hacer de forma manual o automáticamente estableciendo horarios de recolección) se actualizaría el DataFlow.

Por otro lado, TDC almacena versiones de los metadatos. Si se activa la opción para poder verlos, seguiríamos teniendo ambas versiones y se podría mostrar a los usuarios mediante la interfaz *Metadata Explorer* aquella versión que nos interesase. Se puede establecer que automáticamente se muestre la última versión o que el cambio tenga que ser manual.

A pesar de la información mostrada, en el menú Manage>Contents podemos ver que hay un *Warning* en el componente SampleTDI que nos informa de problemas de conexión, esto aparece al no haber unido este Modelo al resto de modelos a los que hace referencia:

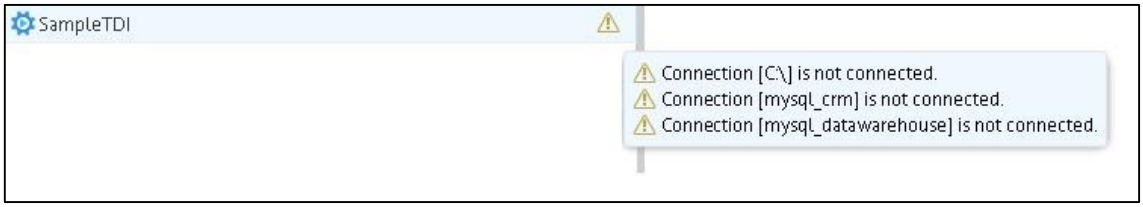


Ilustración 42. DataFlow del trabajo loadDimensions

La solución a este *Warning* se muestra en la siguiente sección.

2.6 STITCHING METADATA

Este proceso debe realizarse tras la recolección de metadatos de algunos componentes cuando existen ambigüedades que no pueden ser resueltas automáticamente. La principal función es crear una relación entre modelos para poder analizar el impacto y el linaje de los datos.

Accedemos a *Browse>ArchitectureDiagram* y vemos que disponemos de un Documento, dos bases de datos y un proceso TDI:

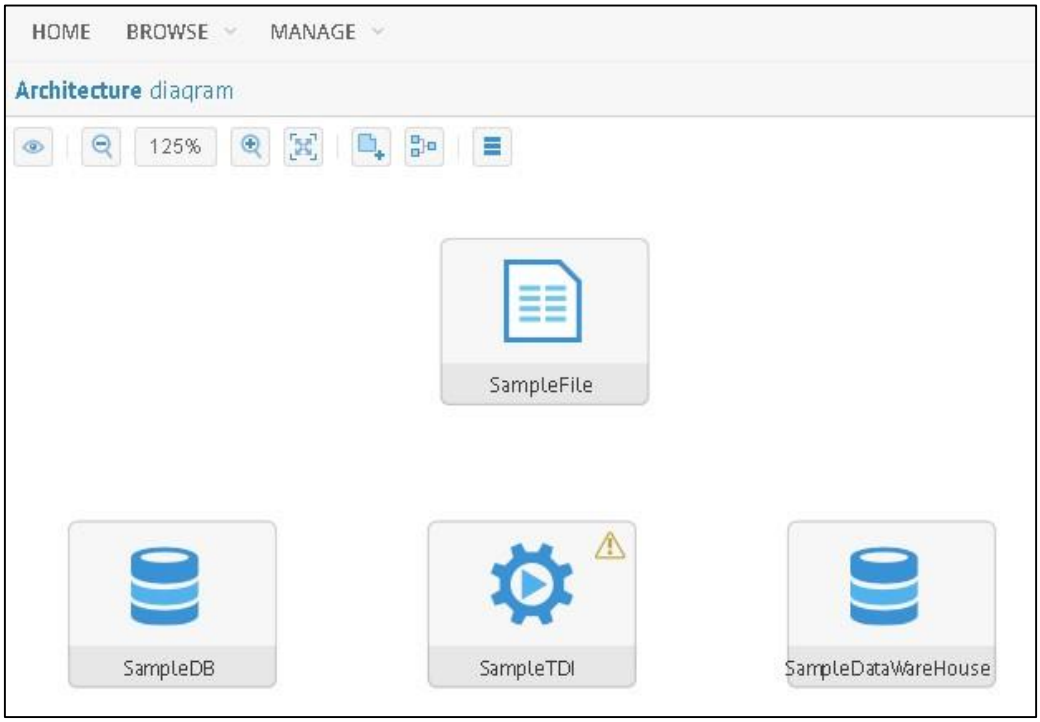


Ilustración 43. Diagrama de arquitectura sin enlazar

De los cuatro componentes disponibles, únicamente el componente SampleTDI tiene un Warning, esto se debe a que dicho componente tiene internamente referencias a otros componentes y no se han establecidos las relaciones. Hacemos click derecho en dicho componente y presionamos en *Edit Connections*:

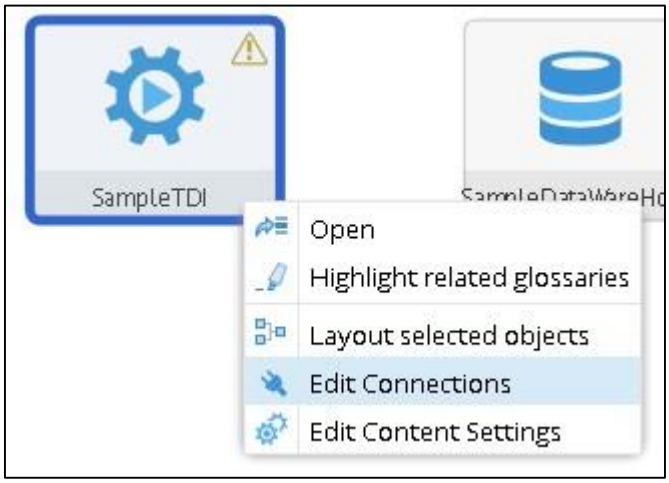


Ilustración 44. Menú de acceso a edición de conexiones

Las conexiones de los componentes se gestionan desde el mismo punto que realizamos las importaciones:

Ilustración 45. Menú de conexiones de Modelo TDI

En la primera fila, en la columna *Store* seleccionamos el fichero:

Ilustración 46. Selección de almacenamiento

Aún con la fila seleccionada, pulsamos en *Stitching Report:*

Ilustración 47. Selección de *Stitching Report*

Desde aquí podemos observar que se ha creado la conexión adecuadamente:

Ilustración 48. Visualización de *Stitching Report* enlazado

Tras realizar el mismo proceso con la segunda fila, vemos en el *Stitching Report* no se ha realizado el enlace como debería:

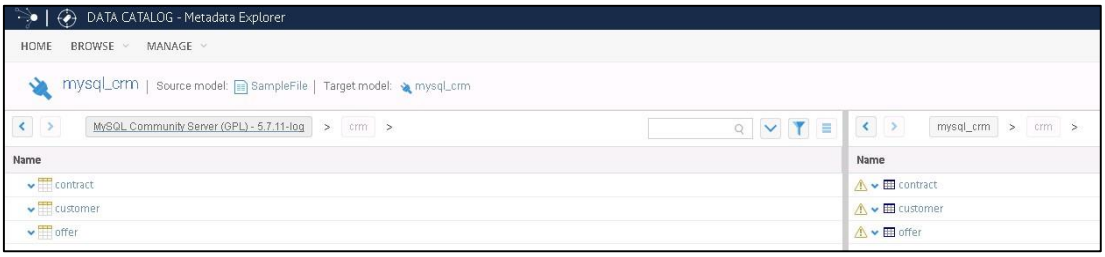


Ilustración 49. Visualización de *Stitching Report* sin enlazar

En este caso tendremos que seleccionar un esquema:

Ilustración 50. Visualización de *Stitching Report* sin enlazar

Dejamos la opción que nos viene por defecto y pulsamos Ok:

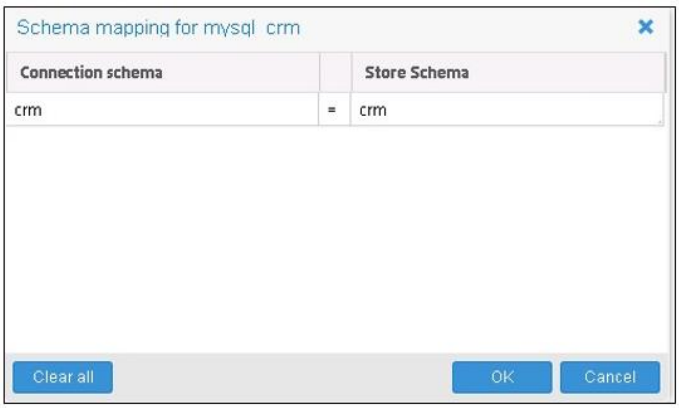


Ilustración 51. Mapeo de esquema

El siguiente paso consiste en realizar el mismo proceso con la tercera fila que corresponde al *datawarehouse*, asignando las columnas *Store* y *Schema/Path:*

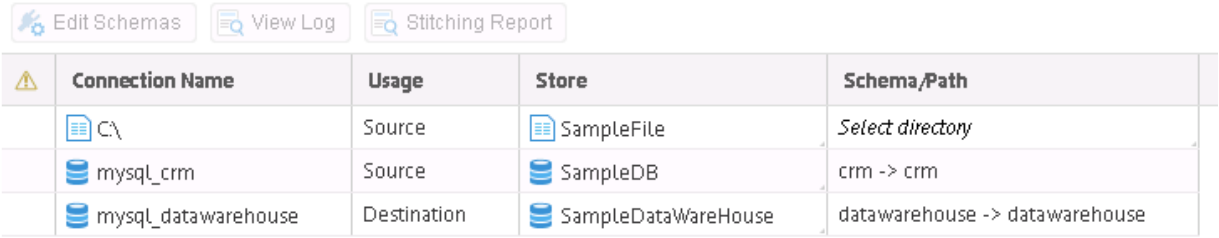


Ilustración 52. Conexiones Preparadas

Tras realizar este proceso es necesario ejecutar una construcción pulsando *Build*.

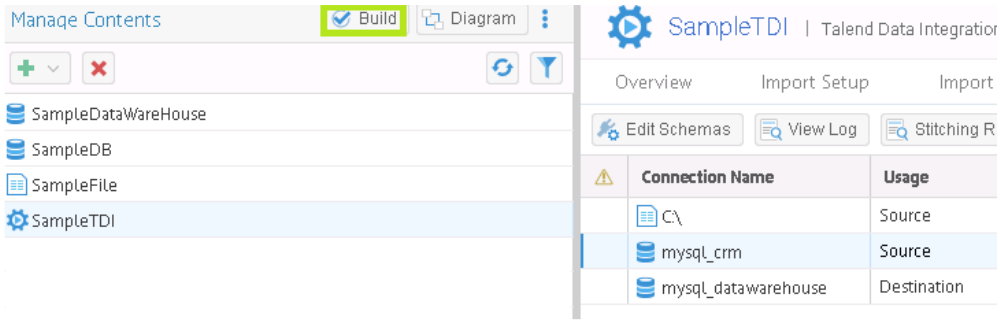


Ilustración 53. Selección de Build

Ahora podemos ver que el *Warning* del componente *SampleTDI* ha desaparecido, por lo que al revisar el *Stitching Report* de las conexiones a bases de datos que acabamos de establecer se muestran los componentes enlazados:

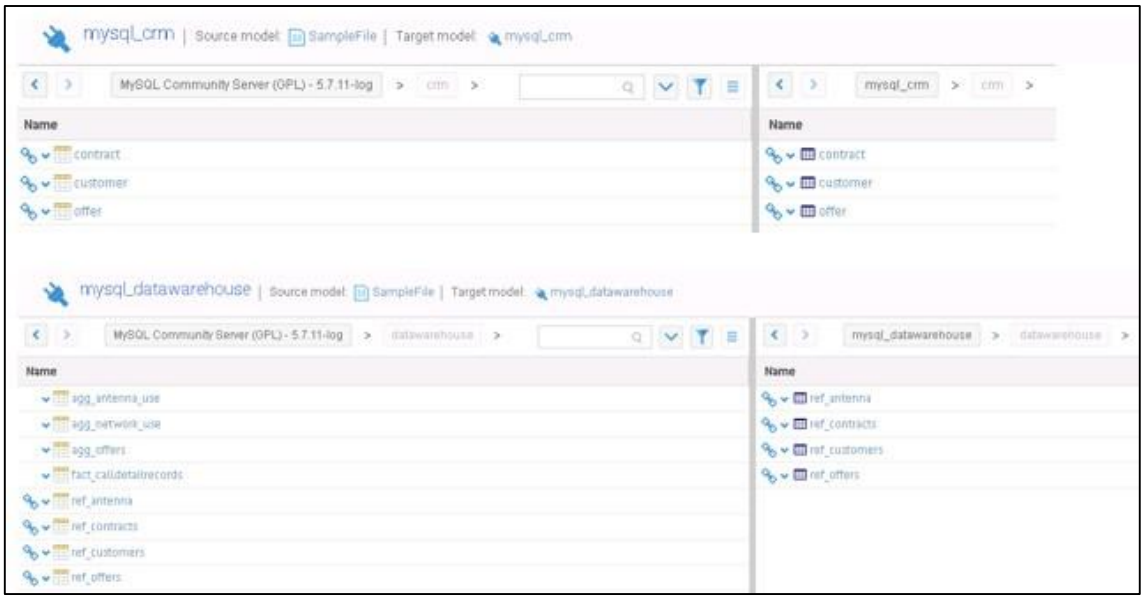


Ilustración 54. Visualización de *Stitching Report* enlazado

Al volver a *Browse>ArchitectureDiagram* apreciamos que los elementos han sido debidamente conectados:

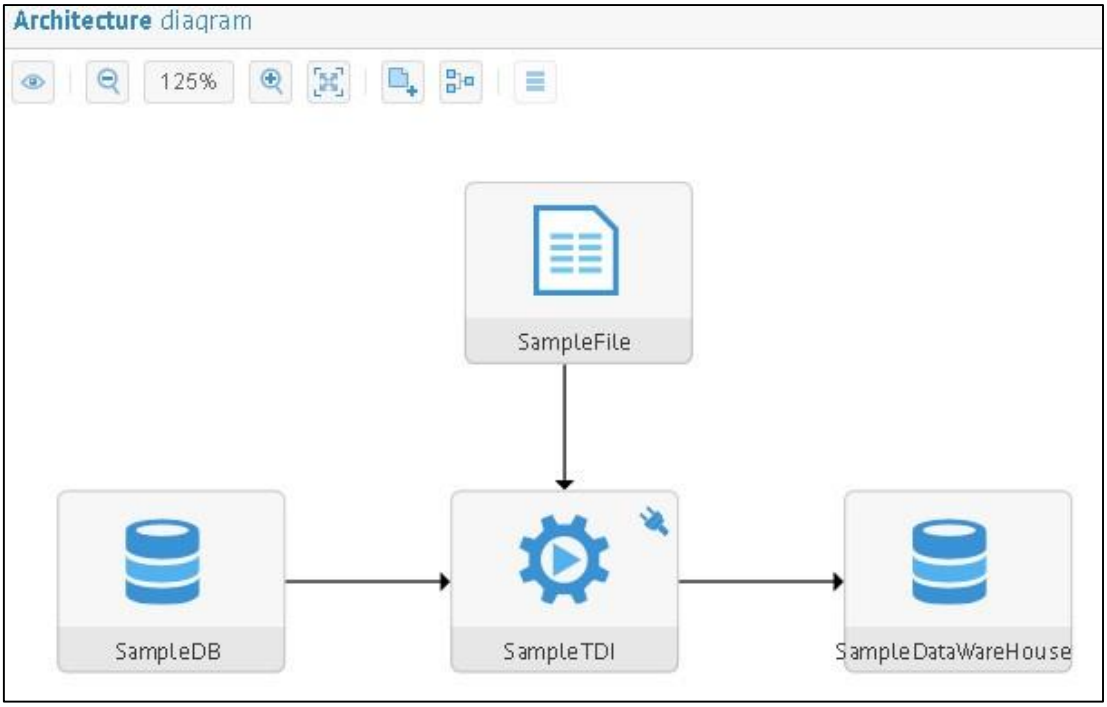


Ilustración 55. Diagrama de arquitectura debidamente enlazado

2.7 CREACIÓN DE GLOSARIO

Para acceder a la creación de un glosario nos dirigimos a *Manage>Contents:*

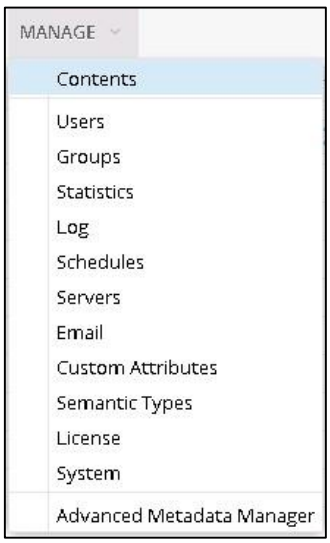


Ilustración 56. Menú de administración de contenido

Añadimos un nuevo glosario a través del símbolo + :

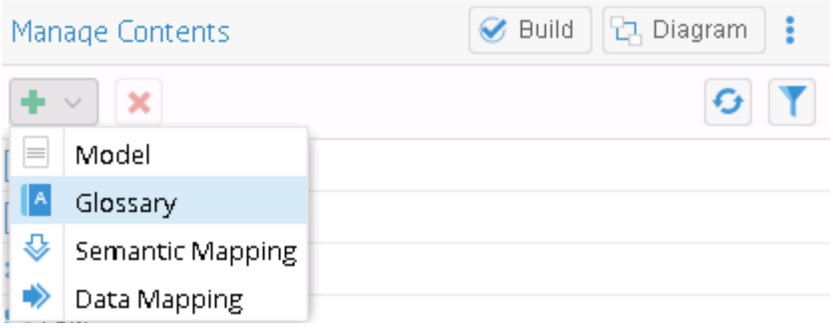


Ilustración 57. Añadir Glosario

Rellenamos los campos de nombre y descripción. Confirmamos pulsando *OK* :

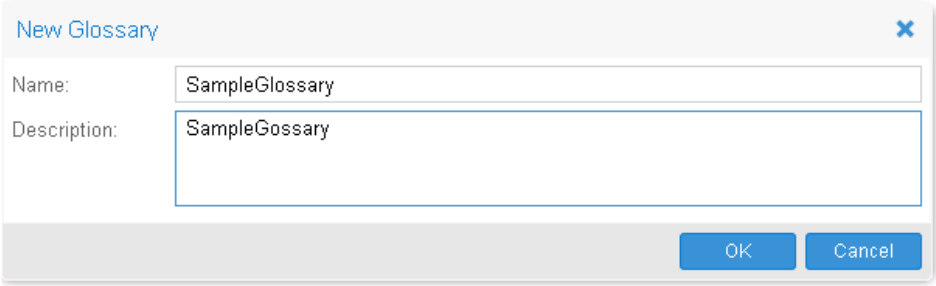


Ilustración 58. Creación de Glosario

A través del menú accedemos a los glosarios que tenemos creados:

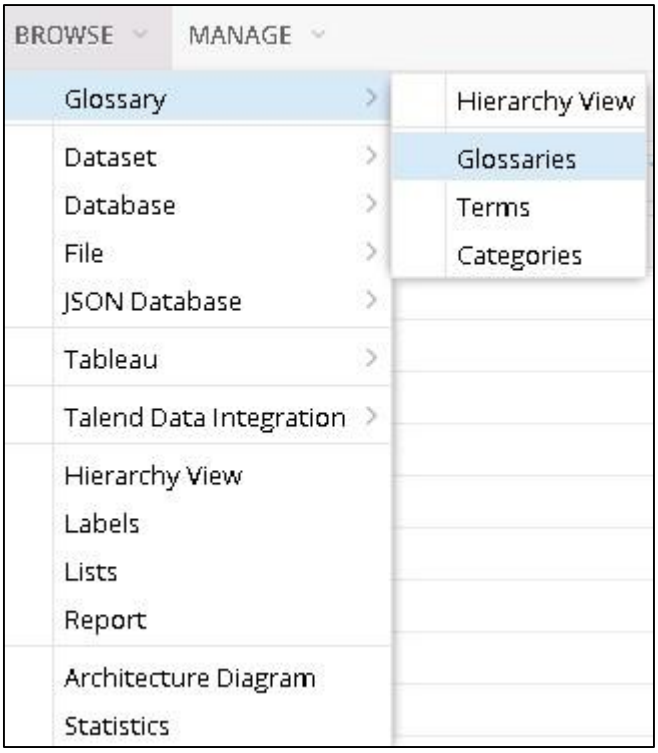


Ilustración 59. Búsqueda de Glosarios

Accedemos al glosario que hemos creado:

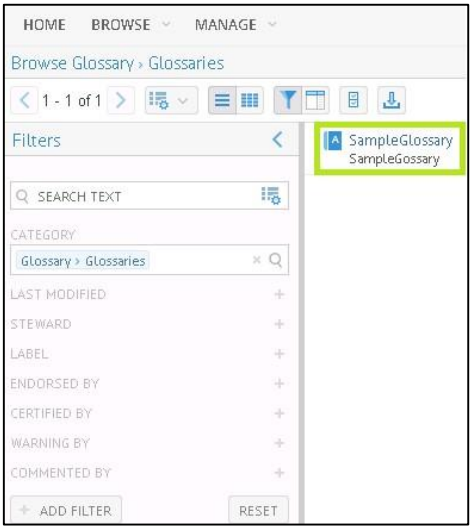


Ilustración 60. Visualización de Glosarios

Desde la visualización previa del glosario, que está vacío, importamos un glosario previamente creado:

Ilustración 61. Menú para la importación de un Glosario

Seleccionamos el archivo y el tipo de delimitador. Pulsamos *Import:*

Ilustración 62. Importación de un Glosario

Aparecerá un log con el progreso del proceso:

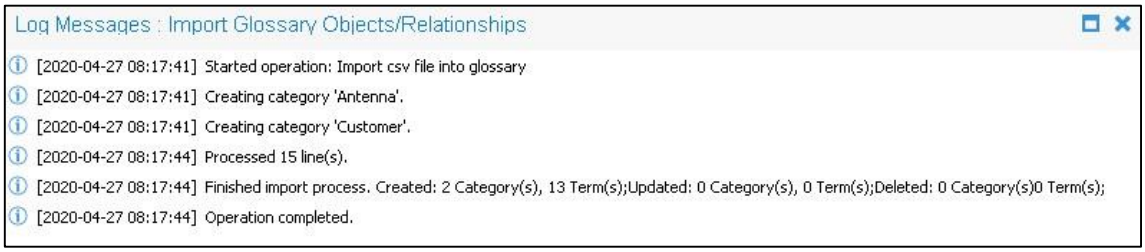


Ilustración 63. Log de importación de un Glosario

Tras finalizar la operación podremos apreciar en la vista previa del Glosario nueva información referente a las categorías y términos importados:

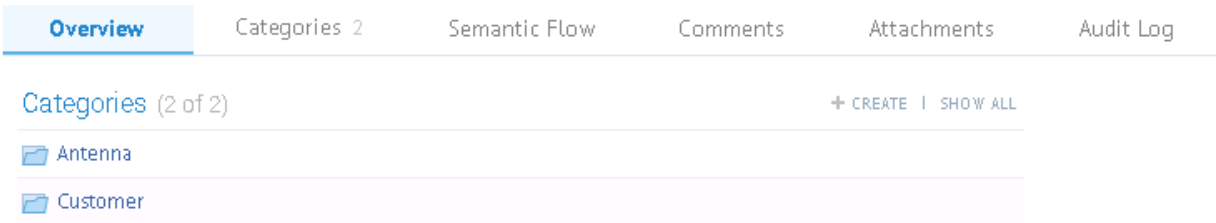


Ilustración 64. Vista previa de Glosario

Siguiendo el mismo proceso de pueden importar relaciones a los glosarios:

Ilustración 65. Importación de Relaciones de un Glosario

Tras acceder a los términos afectados, podemos apreciar que se han creado las relaciones:

Ilustración 66. Relaciones de términos de glosario

En el menú *Browse> Architecture Diagram* vemos el glosario creado, sin relacionar:

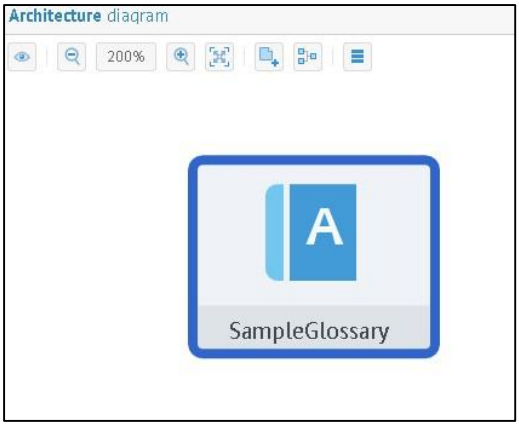


Ilustración 67. Diagrama de Arquitectura de un Glosario desenlazado

El siguiente paso consiste en mapear las relaciones del glosario a los componentes afectados. Para ello, desde el menú de administración de contenidos crearemos un *Semantic Mapping*:

Ilustración 68. Menú de creación de mapa semántico

Rellenamos los campos solicitados. Para seleccionar componentes pulsamos en la lupa:

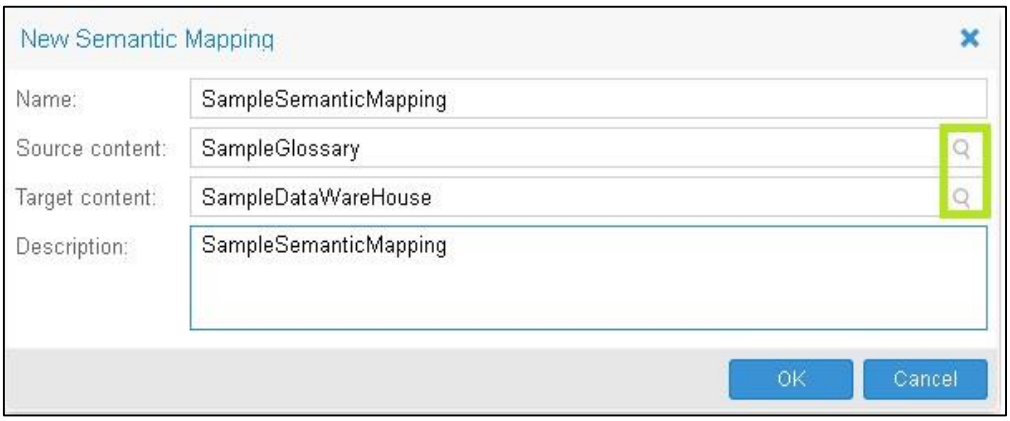


Ilustración 69. Creación de mapa semántico

El diagrama de arquitectura se ha actualizado mostrando la relación entre el glosario y la base de datos datawarehouse mediante el mapa semántico:

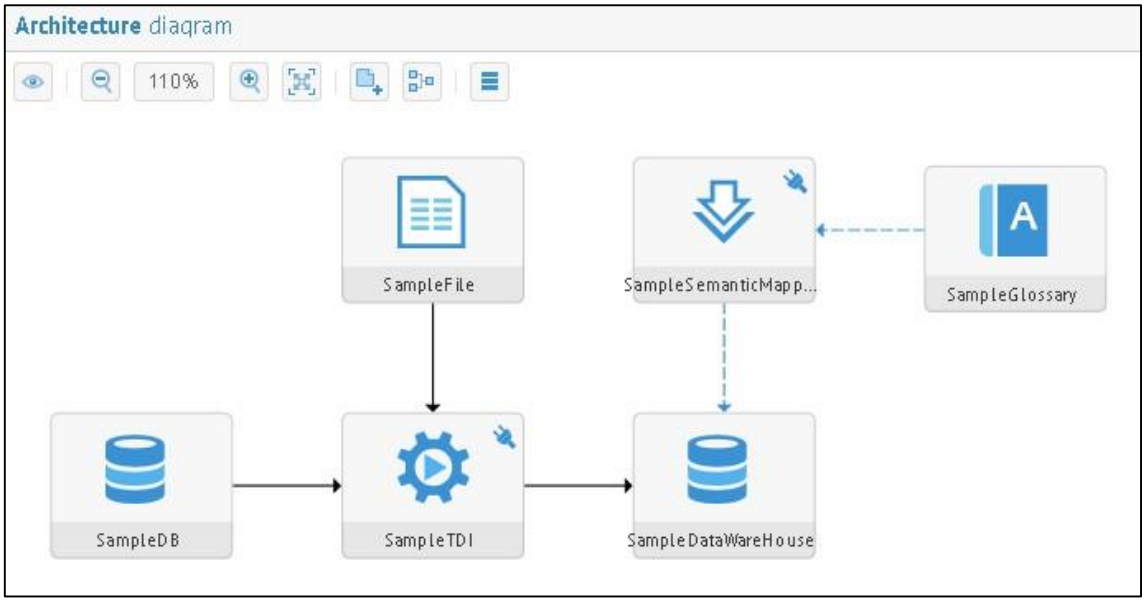


Ilustración 70. Diagrama de arquitectura Glosario relacionado

A continuación, procedemos a realizar un mapeo en profundidad del glosario. Accedemos a *Manage>Advanced Metadata Manager*. Hacemos doble clic en *SampleSemanticMapping* para abrir el componente:

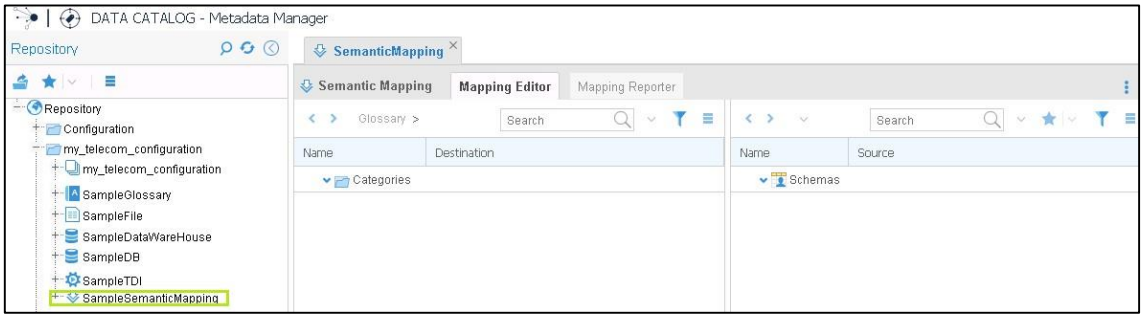


Ilustración 71. Diagrama de arquitectura Glosario relacionado

Expandimos ambas columnas. *Categories>Customer>Terms.*

Schemas>datawarehouse>Tables>ref_customers>Columns. Arrastramos cada elemento de una columna con su correspondiente de la otra columna:

Name	Destination
Address 1	
Address 2	
City	
First Name	
Full Address	
Full Name	
Last Name	
Zipcode	

Name
address1
address2
city
email
firstname
gender
idcustomer
lastname
msisdn
subscriptiondate
tenure
zipcode

Ilustración 72. Mapa Semántico sin mapear

Name	Destination
Address 1	address1
Address 2	address2
City	city
First Name	firstname
Full Address	
Full Name	
Last Name	lastname
Zipcode	zipcode

Name	Source
address1	Address 1
address2	Address 2
city	City
email	
firstname	First Name
gender	
idcustomer	
lastname	Last Name
msisdn	
subscriptiondate	
tenure	
zipcode	Zipcode

Ilustración 73. Mapa Semántico mapeado

Retrocedemos y realizamos el mismo proceso con *Antena* y con *ref_antenna*:

Name	Destination
Area	
City	
Full Coordin...	
X-Coordinates	
Y-Coordinates	

Name
area
city
commission_dt
tech_id
x_coord
y_coord

Ilustración 74. Mapa Semántico sin mapear

Name	Destination
Area	area
City	city
Full Coordinates	
X-Coordinates	x_coord
Y-Coordinates	y_coord

Name	Source
area	Area
city	City
commission_dt	
tech_id	
x_coord	X-Coordinates
y_coord	Y-Coordinates

Ilustración 75. Mapa Semántico mapeado

Cerramos la pestaña *Advanced Metadata Manager* del navegador, y desde el diagrama de arquitectura realizamos doble clic en el componente del mapa semántico. Desde el componente, si accedemos al campo *FirstName* del glosario (columna de la izquierda) podremos visualizar la información referente a dicho término:

SampleSemanticMapping | Source model: SampleGlossary | Target model: SampleDataWareHouse — Import — Export

Show only broken links

Source	Source Path	Target	Target Path
First Name	/SampleGlossary/Categories/Customer/Terms/First Name	firstname	/MySQL Community Server (GPL) - 5.7.11-log /Schemas/datawarehouse/Tables/r...
Last Name	/SampleGlossary/Categories/Customer/Terms/Last Name	lastname	/MySQL Community Server (GPL) - 5.7.11-log /Schemas/datawarehouse/Tables/r...
Address 1	/SampleGlossary/Categories/Customer/Terms/Address 1	address1	/MySQL Community Server (GPL) - 5.7.11-log /Schemas/datawarehouse/Tables/r...
Address 2	/SampleGlossary/Categories/Customer/Terms/Address 2	address2	/MySQL Community Server (GPL) - 5.7.11-log /Schemas/datawarehouse/Tables/r...
Zipcode	/SampleGlossary/Categories/Customer/Terms/Zipcode	zipcode	/MySQL Community Server (GPL) - 5.7.11-log /Schemas/datawarehouse/Tables/r...
City	/SampleGlossary/Categories/Customer/Terms/City	city	/MySQL Community Server (GPL) - 5.7.11-log /Schemas/datawarehouse/Tables/r...
Area	/SampleGlossary/Categories/Antenna/Terms/Area	area	/MySQL Community Server (GPL) - 5.7.11-log /Schemas/datawarehouse/Tables/r...
City	/SampleGlossary/Categories/Antenna/Terms/City	city	/MySQL Community Server (GPL) - 5.7.11-log /Schemas/datawarehouse/Tables/r...
X-Coordinates	/SampleGlossary/Categories/Antenna/Terms/X-Coordinates	x_coord	/MySQL Community Server (GPL) - 5.7.11-log /Schemas/datawarehouse/Tables/r...
Y-Coordinates	/SampleGlossary/Categories/Antenna/Terms/Y-Coordinates	y_coord	/MySQL Community Server (GPL) - 5.7.11-log /Schemas/datawarehouse/Tables/r...

Ilustración 76. Vista previa del componente Mapa Semántico

Desde la pestaña *Semantic Flow*, con la opción de *Diagram* y el tipo *Usage* podemos ver el diagrama de uso de dicho término:

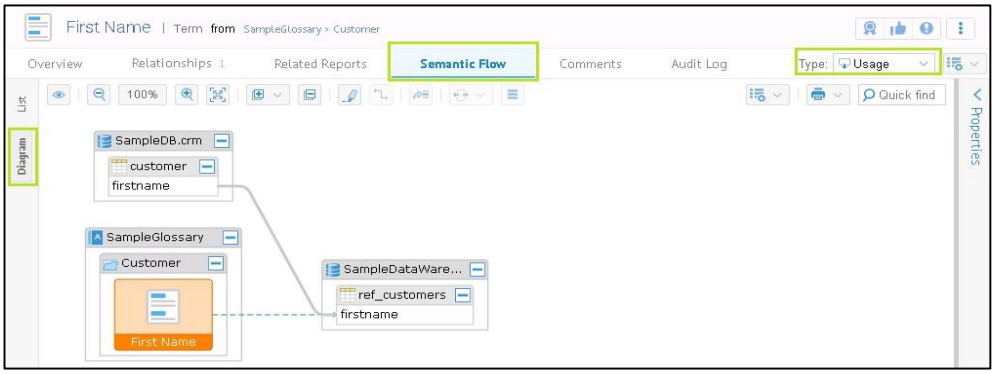


Ilustración 77. Diagrama de uso de un término de un Mapa Semántico

2.8 TIPOS SEMÁNTICOS

Para administrar los tipos semánticos accedemos al menú *Manage>Semantic Types:*

Ilustración 78. Diagrama de uso de un término de un Mapa Semántico

Pulsando en *Import* podemos buscar e importar un Tipo Semántico previamente creado:

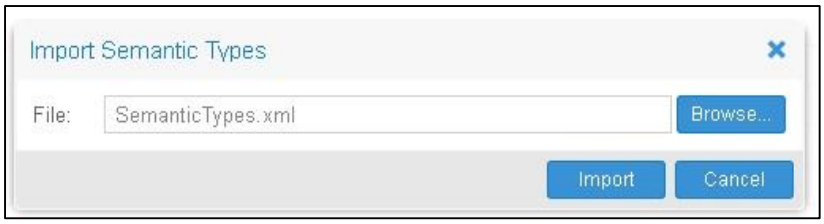


Ilustración 79. Importación de Tipo Semántico

Tras la importación comprobaríamos que se ha importado correctamente y lo publicamos:

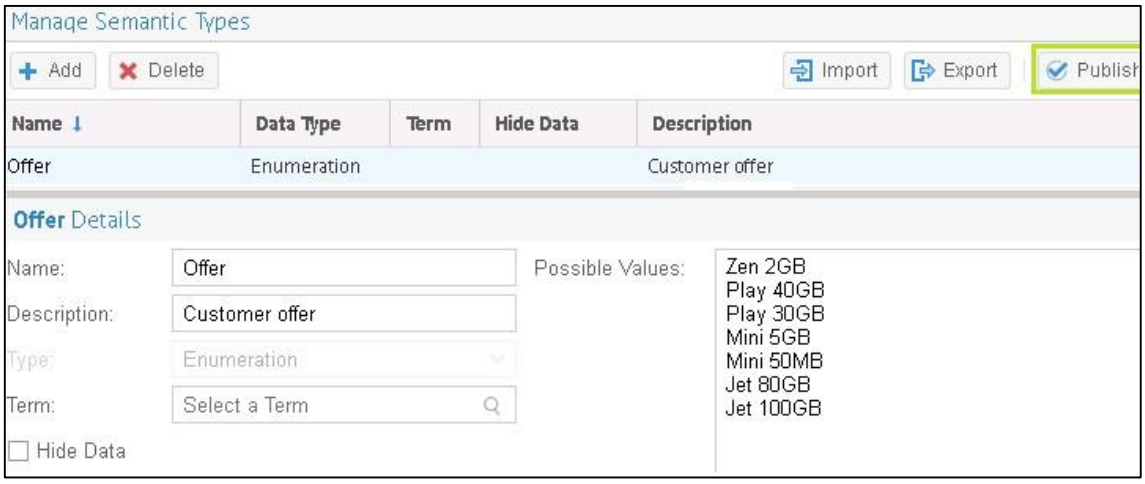


Ilustración 80. Publicación de un Tipo Semántico

Desde el buscador de la esquina superior derecha accedemos a la comuna *offername* donde aplicaremos el Tipo Semántico *offer* importado:

Ilustración 81. Resultado de búsqueda *offername column*

Se puede apreciar que la columna *offername* aún no tiene ningún Tipo Semántico asignado, abrimos el menú pulsando en los tres puntos y hacemos clic en *Refresh profiling data*:

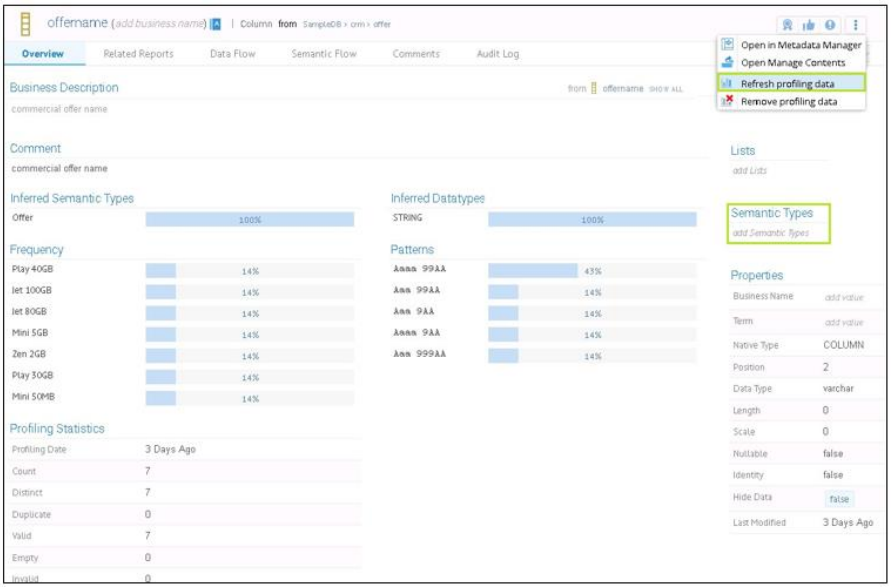


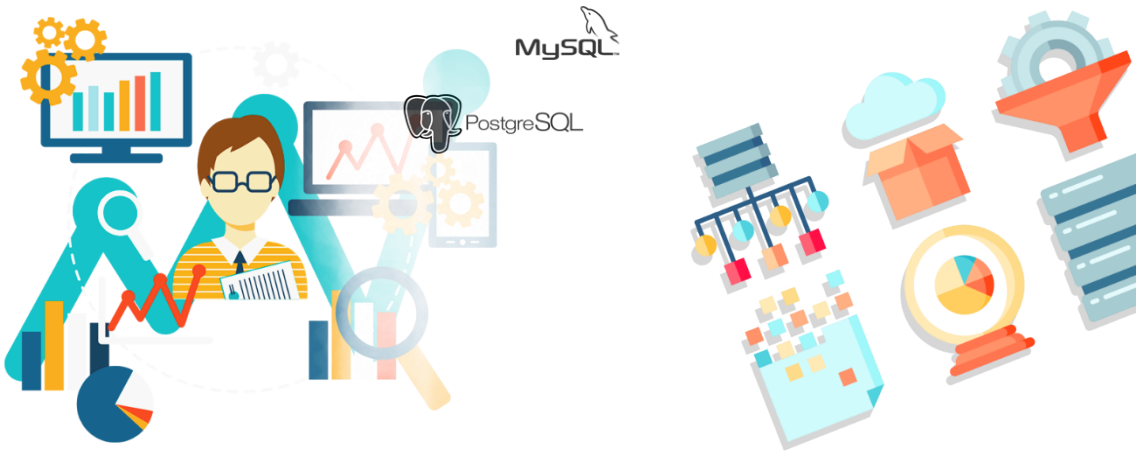
Ilustración 82. Previsualización *offername*

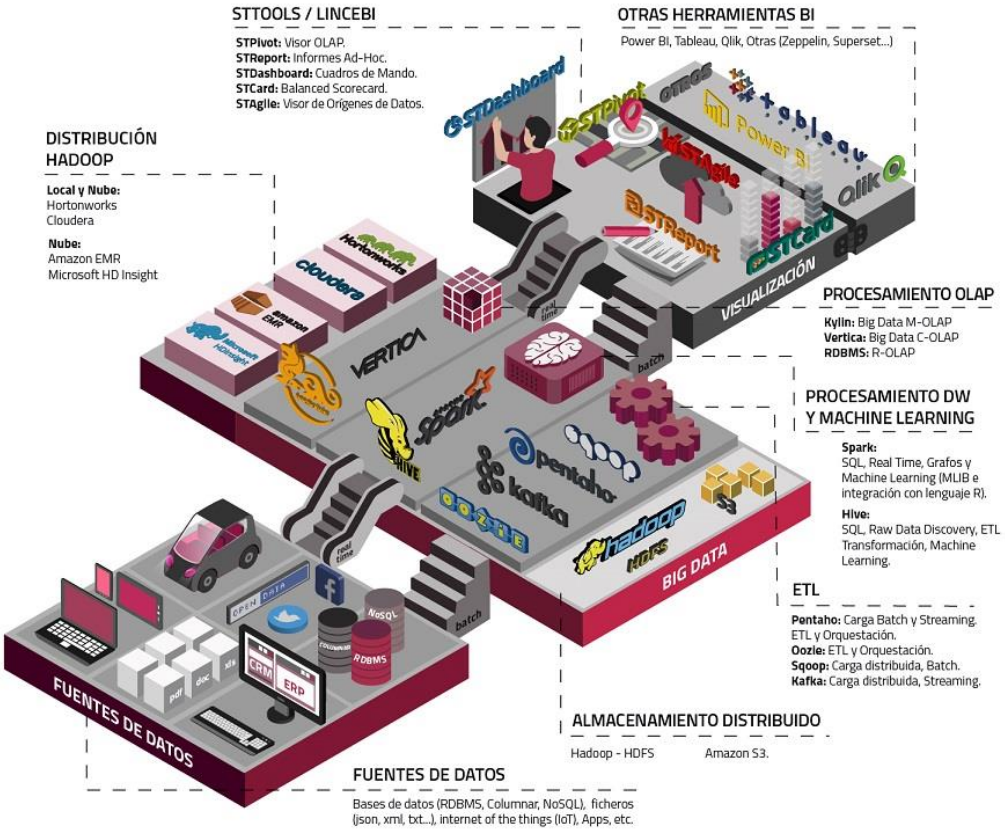
Dejando las opciones predeterminada pulsamos en *OK.* Nos parecerá un log informando del progreso del proceso. Tras terminar podemos apreciar que se ha detectado automáticamente el Tipo Semántico:

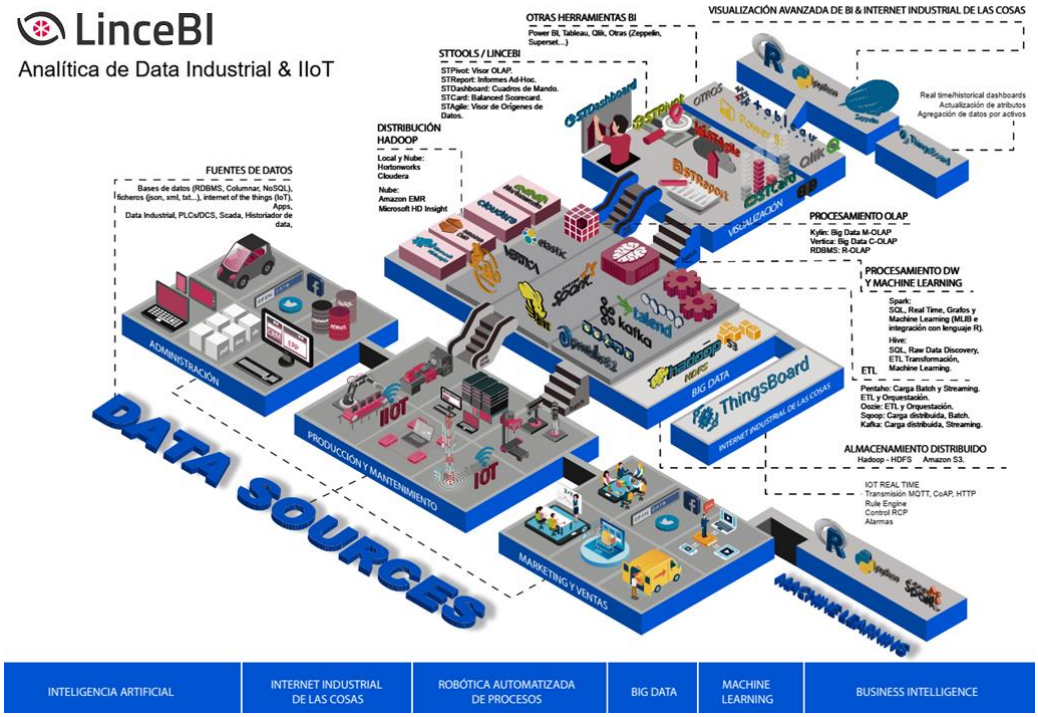
Ilustración 83. Tipo Semántico detectado

TECNOLOGÍAS

Trabajamos con las principales tecnologías y somos Partners Certificados de Vertica, Talend, Microsoft, Snowflake, Kylligence, Pentaho, etc.




STTOOLS / LINCEBI
STPivot: Visor OLAP.
STReport: Informes Ad-Hoc.
STDashboard: Cuadros de Mando.
STCard: Balanced Scorecard.
STAgile: Visor de Orígenes de Datos.
OTRAS HERRAMIENTAS BI
Power BI, Tableau, Qlik, Otras (Zeppelin, Superset...)
DISTRIBUCIÓN HADOOP
Local y Nube:
Hortonworks
Cloudera
Nube:
Amazon EMR
Microsoft HD Insight
VISUALIZACIÓN
PROCESAMIENTO OLAP
Kylin: Big Data M-OLAP
Vertica: Big Data C-OLAP
RDBMS: R-OLAP
PROCESAMIENTO DW Y MACHINE LEARNING
Spark:
SQL, Real Time, Grafos y Machine Learning (MLIB e integración con lenguaje R).
Hive:
SQL, Raw Data Discovery, ETL Transformación, Machine Learning.
BIG DATA
ETL
Pentaho: Carga Batch y Streaming. ETL y Orquestación.
Oozie: ETL y Orquestación.
Sqoop: Carga distribuida, Batch.
Kafka: Carga distribuida, Streaming.
FUENTES DE DATOS
ALMACENAMIENTO DISTRIBUIDO
Hadoop - HDFS
Amazon S3.
FUENTES DE DATOS
Bases de datos (RDBMS, Columnar, NoSQL), ficheros (json, xml, txt...), internet of the things (IoT), Apps, etc.


LinceBI
Analítica de Data Industrial & IIoT
OTRAS HERRAMIENTAS BI
VISUALIZACIÓN AVANZADA DE BI & INTERNET INDUSTRIAL DE LAS COSAS
STTOOLS / LINCEBI
DISTRIBUCIÓN HADOOP
FUENTES DE DATOS
PROCESAMIENTO OLAP
PROCESAMIENTO DW Y MACHINE LEARNING
ETL
ALMACENAMIENTO DISTRIBUIDO
DATA SOURCES
ADMINISTRACIÓN
PRODUCCIÓN Y MANTENIMIENTO
MARKETING Y VENTAS
MACHINE LEARNING
INTELIGENCIA ARTIFICIAL
INTERNET INDUSTRIAL DE LAS COSAS
ROBÓTICA AUTOMATIZADA DE PROCESOS
BIG DATA
MACHINE LEARNING
BUSINESS INTELLIGENCE

INFORMACIÓN SOBRE STRATEBI

Stratebi es una empresa española, con sede en Madrid y oficinas en Barcelona, Alicante y Sevilla, creada por un grupo de profesionales con amplia experiencia en sistemas de información, soluciones tecnológicas y procesos relacionados con soluciones de Open Source y de inteligencia de Negocio.

Esta experiencia, adquirida durante la participación en proyectos estratégicos en compañías de reconocido prestigio a nivel internacional, se ha puesto a disposición de nuestros clientes.

Somos **Partners Certificados en Microsoft PowerBI** con una dilatada experiencia

Stratebi es la única empresa española que ha estado presente todos los Pentaho Developers celebrados en Europa habiendo organizado el de España.

En Stratebi nos planteamos como **objetivo** dotar a las compañías e instituciones, de herramientas escalables y adaptadas a sus necesidades, que conformen una estrategia Business Intelligence capaz de rentabilizar la información disponible. Para ello, nos basamos en el desarrollo de soluciones de Inteligencia de Negocio, mediante tecnología Open Source.

Stratebi **son** profesores y responsables de proyectos **del Master en Business Intelligence de la Universidad UOC, UCAM, EOI…**

Los profesionales de Stratebi son los creadores y autores del primer weblog en español sobre el mundo del Business Intelligence, Data Warehouse, CRM, Dashboards, Scorecard y Open Source. Todobi.com

Stratebi es partner de las principales soluciones Analytics: Microsoft Power BI, Talend, Pentaho, Vertica, Snowflake, Kyligence, Cloudera…

Todo Bi, se ha convertido en una referencia para el conocimiento y divulgación del Business Intelligence en español.

OTROS

Trabajamos en los principales sectores y con algunas de las compañías y organizaciones más importantes de España.

SECTOR PRIVADO

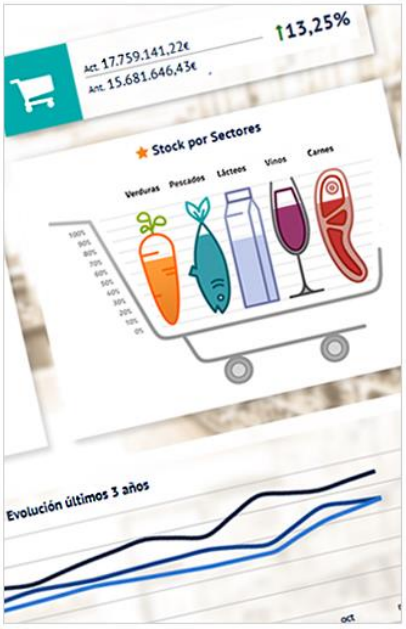
GLOBALIA
ITALFARMACO
Unilever
pepecar.com
20:20mobile
EROSKI
amper
educaweb (*)
EQUIFAX
BBVA
Blanco
PROSEGUR
FEMUR
Telefónica
FERMAX
BOSCH
AMNESTY INTERNATIONAL

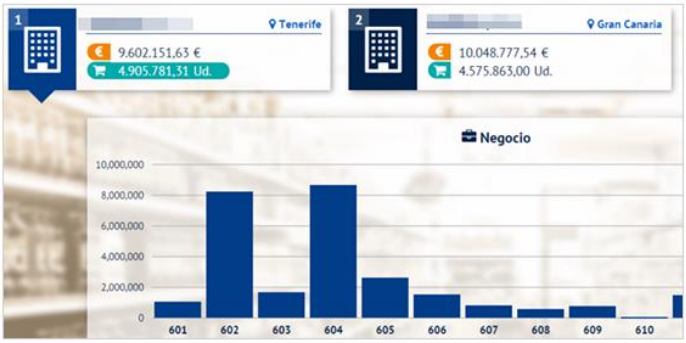
SECTOR PÚBLICO

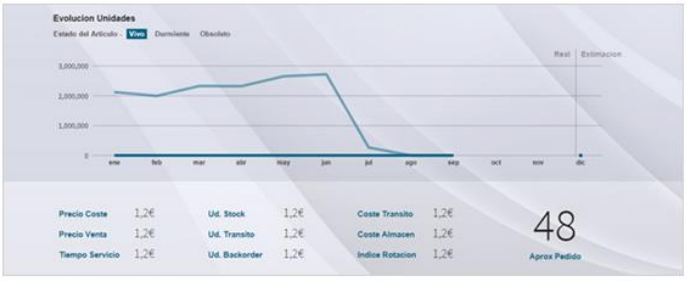
Gobierno de Canarias
Consejería de Educación, Universidades y Sostenibilidad
UAM UNIVERSIDAD AUTONOMA DE MADRID
Caixa d'Enginyers
Diputació Tarragona
Ayuntamiento de Logroño
CSIC
Universitat de Lleida
IDI INSTITUT DE DIAGNÒSTIC PER LA IMATGE
Ayuntamiento de Arrecife
Gobierno de La Rioja
Dipsalut
Organisme Autònom de Salut Pública
Diputació de Girona
Universidad Zaragoza
AJUNTAMENT DE REUS
SaludMadrid
EOI Escuela de organización industrial

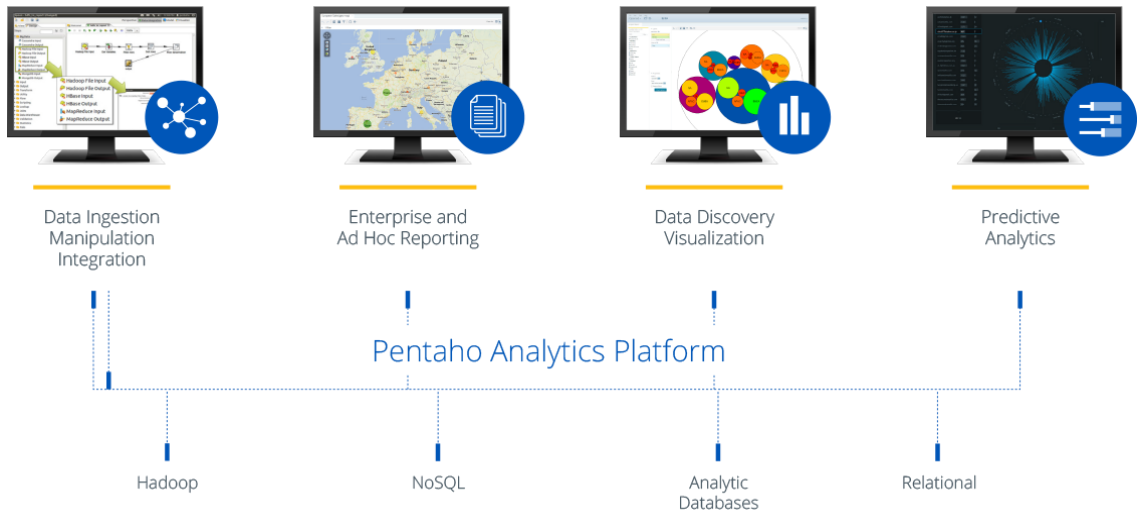
EJEMPLOS DE DESARROLLOS ANALYTICS

A continuación, se presentan **ejemplos de algunos screenshots** de cuadros de mando diseñados por Stratebi, con el fin de dar a conocer lo que se puede llegar a obtener, así como Demos Online en la web de Stratebi:



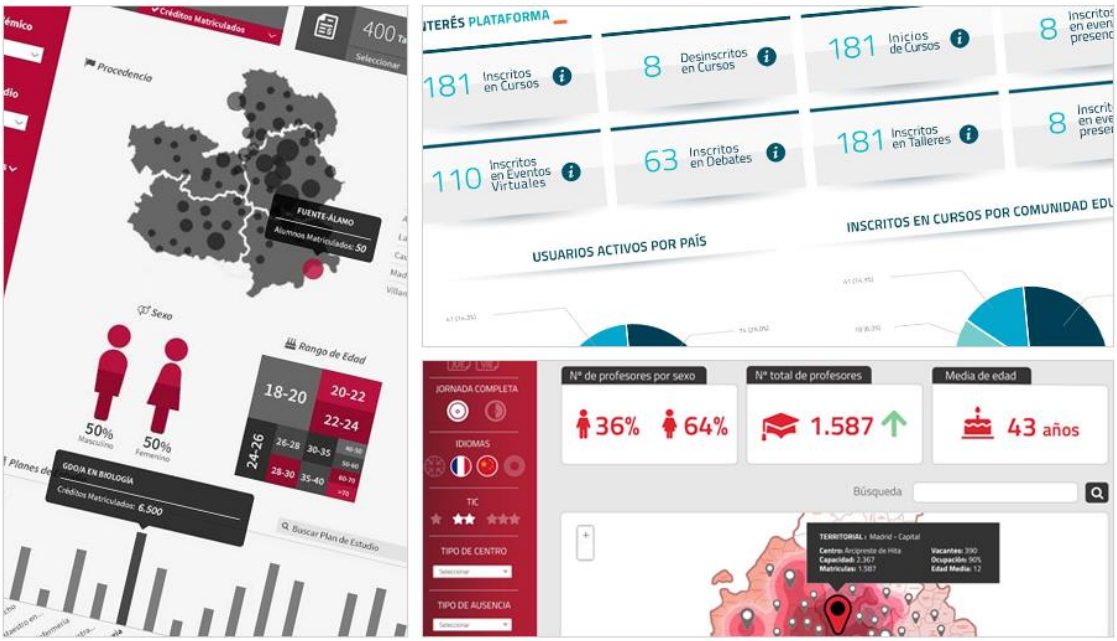


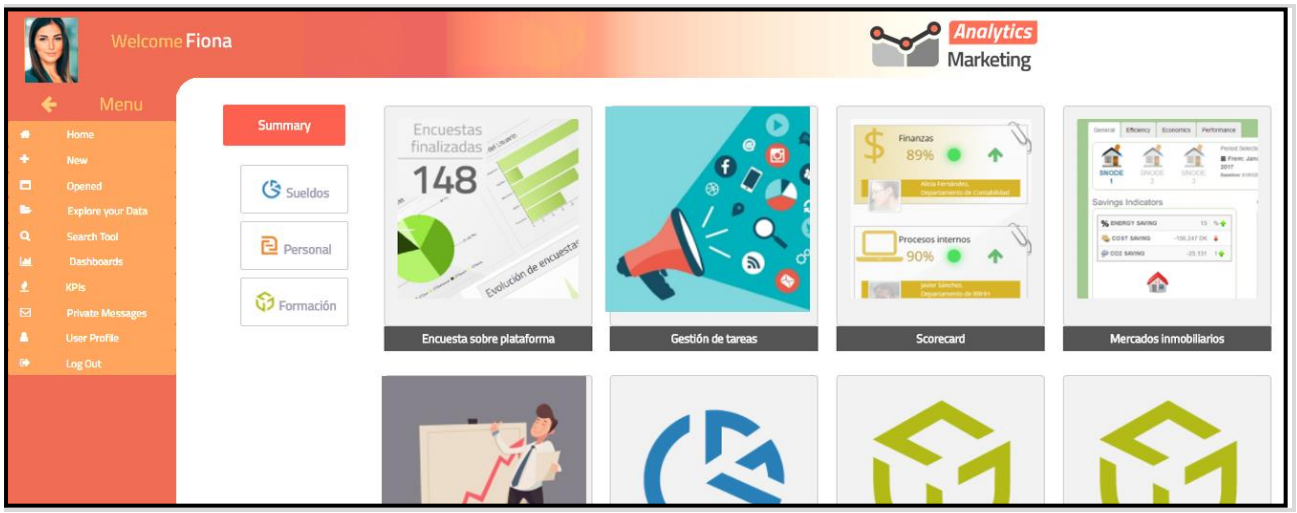





facebook DASHBOARD COMENTARIOS
Sexo
Sentimiento
Localización
65%
35%
28%
60%
12%
Tipo de recurso
Fotos
Status
Vídeos
Preguntas
Evolución Comentarios por días


smart MADRID
Último Actualización
18:24
TOP 5 PUNTOS DE MEDICIÓN
PROXIMIDAD DE PUNTOS DE MEDICIÓN
EVOLUCIÓN INTENSIDAD DE TRÁFICO


Procedencia
Sexo
Rango de Edad
50%
50%
18-20
20-22
22-24
24-26
26-28
28-30
30-35
35-40
181 Inscritos en Cursos
8 Desinscritos en Cursos
181 Inicios de Cursos
110 Inscritos en Eventos Virtuales
63 Inscritos en Debates
181 Inscritos en Talleres
USUARIOS ACTIVOS POR PAÍS
INSCRITOS EN CURSOS POR COMUNIDAD EDU
JORNADA COMPLETA
IDIOMAS
TIC
TIPO DE CENTRO
TIPO DE AUSENCIA
Nº de profesores por sexo
Nº total de profesores
Media de edad
36%
64%
1.587
43 años
Búsqueda


Welcome Fiona
Analytics
Marketing
Menu
Home
New
Opened
Explore your Data
Search Tool
Dashboards
KPIs
Private Messages
User Profile
Log Out
Summary
Sueldos
Personal
Formación
Encuestas finalizadas
148
Finanzas
89%
Procesos internos
90%
Savings Indicators
Encuesta sobre plataforma
Gestión de tareas
Scorecard
Mercados inmobiliarios