



Talend Master Data Management (MDM)

Conceptos principales



Conceptos principales

Índice

- 1) Qué es Master Data Management (MDM)
- 2) Governance (Gestión)
- 3) Subject Area (Area de Negocio)
- 4) Master Data (Datos Maestros)
- 5) Arquitectura de MDM
- 6) MDM Governance (Gestión)
- 7) Data Quality (Calidad de Datos) y MDM
- 8) Roles y Responsabilidades de MDM
- 9) Acciones de MDM

Master Data Management (MDM)



¿Qué es Master Data Management (MDM)?

- Master Data Management es una disciplina esencial para obtener una sola visión coherente de las principales entidades empresariales de una empresa: productos, proveedores, empleados y otros.
- Las soluciones MDM permiten la sincronización de datos maestros en toda la empresa.



Governance

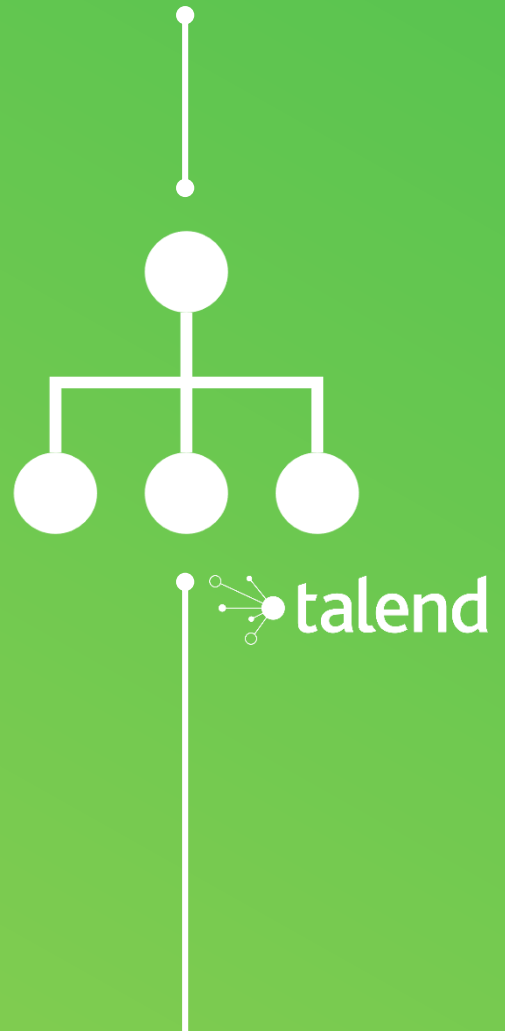


Governance (Gestión)

- MDM en el proceso de *governance* (gestión) facilita la creación del *master data*.
- El objetivo del MDM es construir el *master data* para cada sistema en la empresa.
- Por ejemplo, se puede dedicar el 80% del esfuerzo en una campaña de publicidad en recopilar la lista de productos a promocionar.
- Con MDM, este esfuerzo se reduciría y se podría emplear en la campaña en sí.

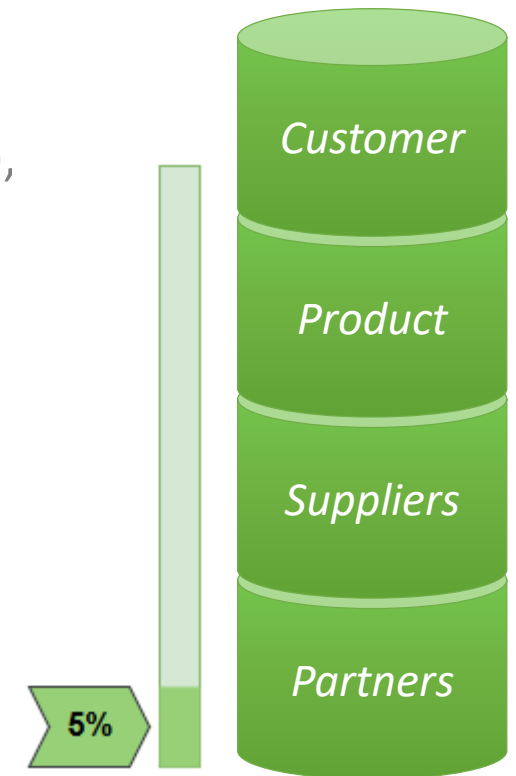
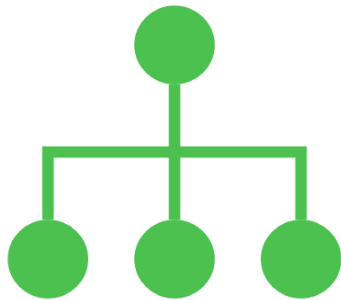


Subject Area



Subject Area (Area de Negocio)

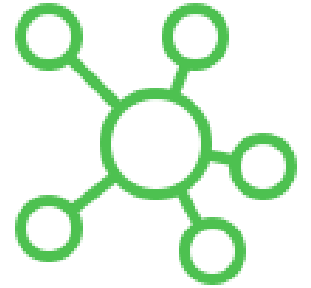
- Master Data representa un pequeño porcentaje, tal vez el 5% del volumen de datos organizativos. Datos de calidad, no datos de cantidad.
- Los datos se suelen organizar en *subject areas* o áreas de negocio.
- Por ejemplo: *Customer* (Cliente), *Product* (Producto), *Suppliers* (Proveedores), *Partners* (Socios)
- MDM es un programa iterativo, implementado en toda la organización a lo largo del tiempo.



Master Data



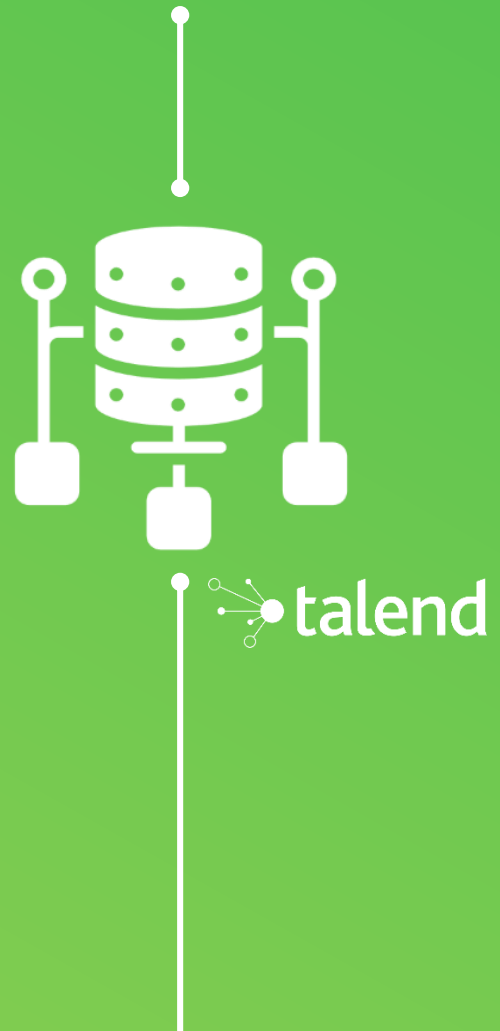
Master Data (Datos Maestros)



- Master Data se trata de *construir una vez, usar a menudo*.
- Datos útiles a las aplicaciones en toda la empresa.
- Se identifican atributos dentro de *Master Data*, como dirección del cliente.
- Los beneficios técnicos tangibles del MDM incluyen (para cada *subject area*):
 - 1. Un modelo de datos multi-aplicación, escalable a toda la empresa
 - 2. Procesos de flujo de trabajo que soportan la generación, verificación, homogeneización de los datos (automatizados y con intervención)
 - 3. Mejora de la calidad de los datos

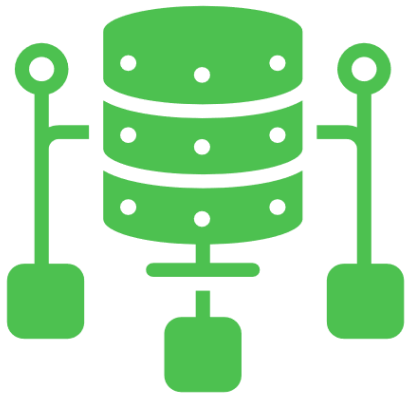


Master Data

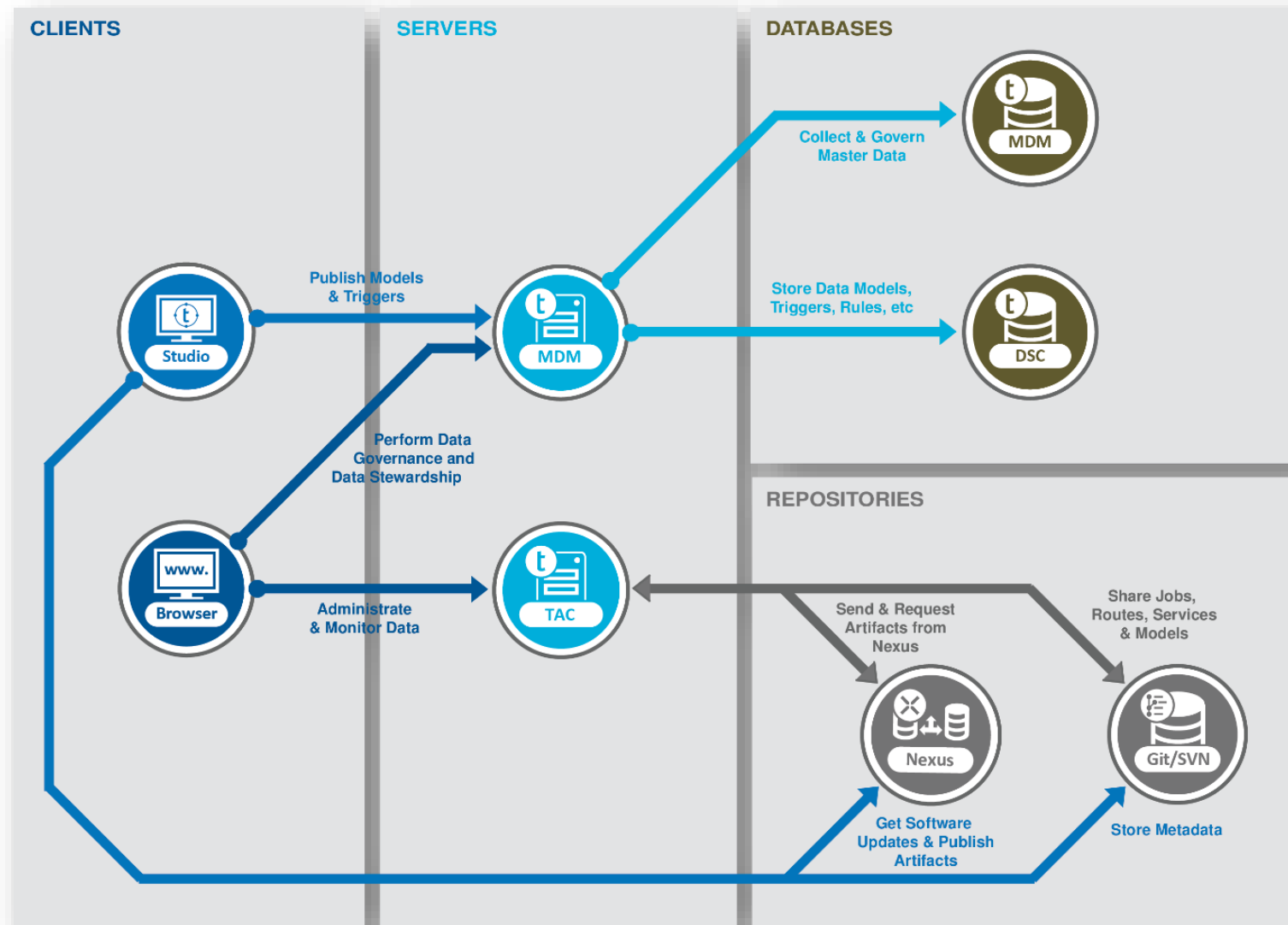


Arquitectura de MDM

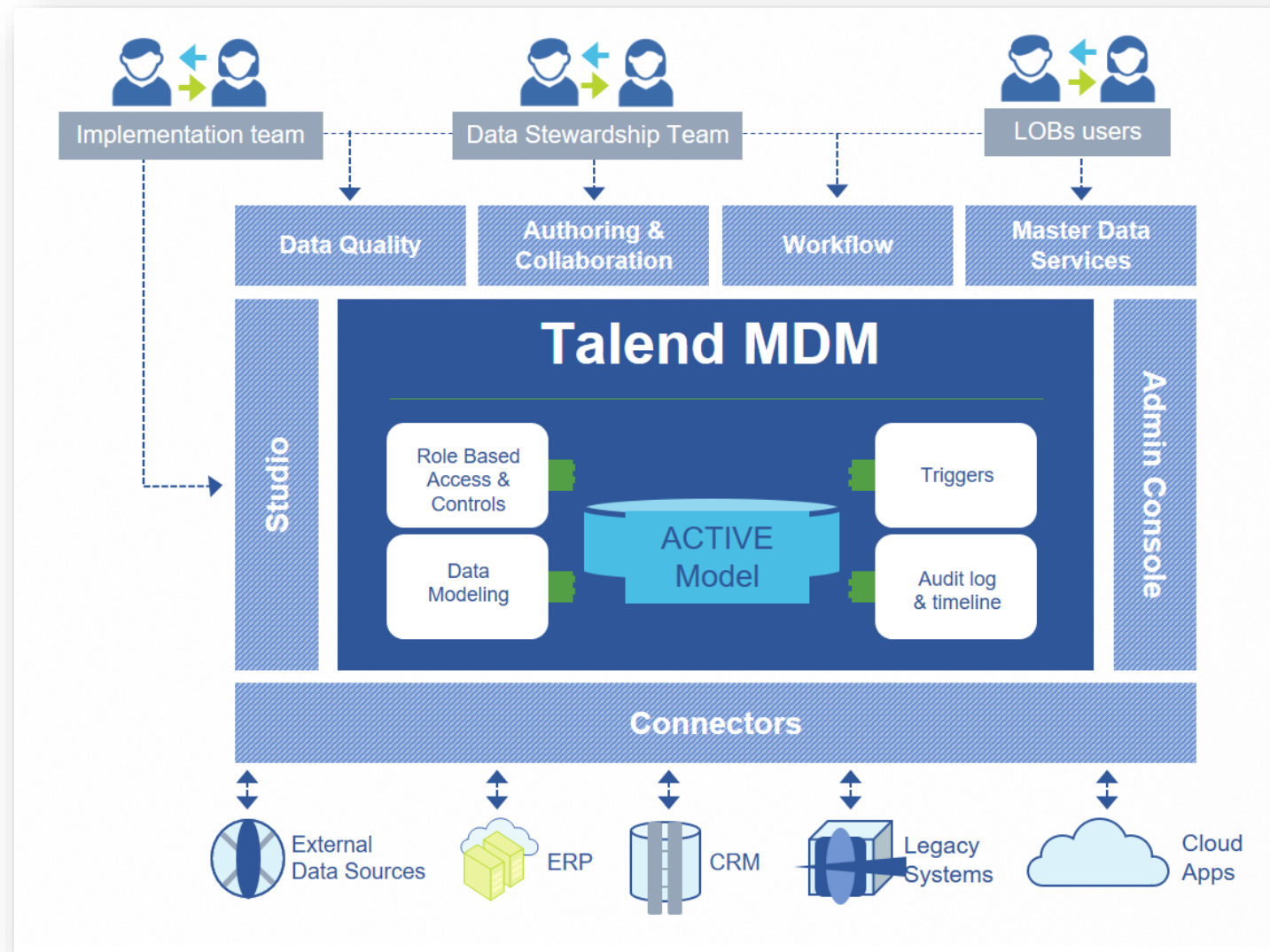
- Un *Data WareHouse* puede ser una pieza clave de la arquitectura, recibiendo master data del MDM.
- Otra estrategia es un *MDM Hub*, donde el dato estaría físicamente en el modelo operaciones y lo que se tendrían son punteros a estos datos.
 - Aquí la calidad del dato dependería de los sistemas de origen.
 - El rendimiento suele ser malo y es útil sólo con un volumen pequeño de datos.
- Otra arquitectura es replicando el *Master Data* en el *Hub*, de modo que se aísla del origen. De este modo, se mejora el rendimiento, limpieza y control de datos.



Arquitectura de MDM (técnica)



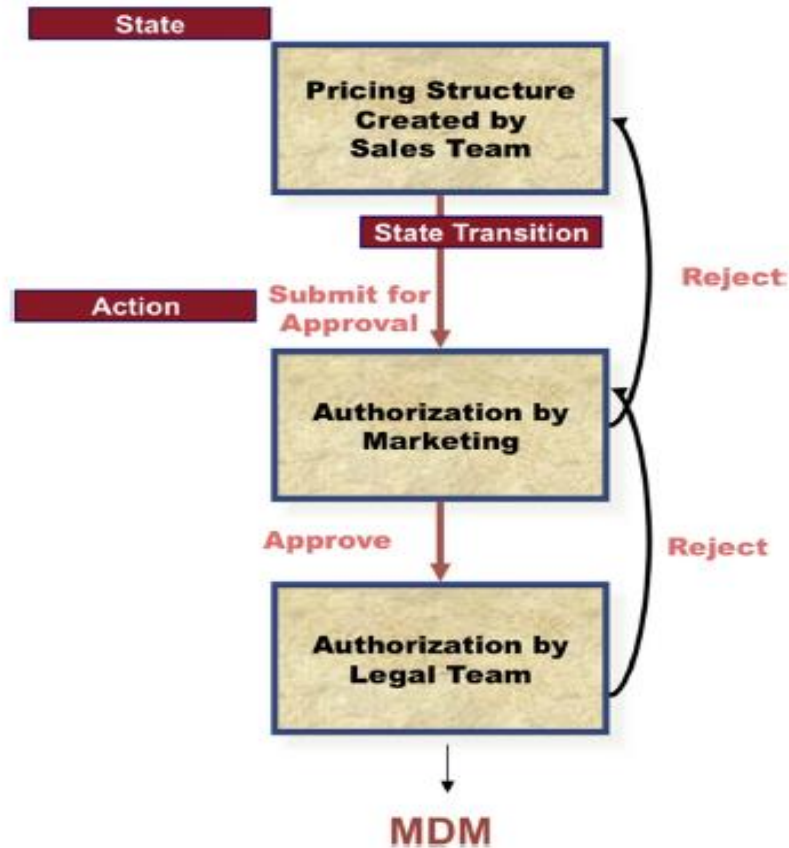
Arquitectura de MDM (funcional)



MDM Governance



MDM Governance (Gestión)



Governance (Gestión) son los procesos involucrados en el desarrollo del 'Master Data'

Por ejemplo, para que se dé de alta un nuevo producto se necesita:

- El coste (Manager de Compras)
- La imagen (Manager de Marketing)
- Garantía (Manager de Servicio)



talend

Data Quality & MDM



Data Quality (Calidad de Datos) y MDM

- Un alto nivel de calidad es uno de los objetivos de MDM.
- *Data Quality* para asegurar la calidad del dato.
- Definir y aplicar reglas de calidad a los datos.



Roles y Responsabilidades de MDM



Roles y Responsabilidades de MDM

- Data stewards: formación del *workflow* y ejecución del mismo.
- Business analysts: analizar las *business* áreas de las que son expertos.
- Project Sponsor: ejecutivo que entiende la importancia de la información y que contribuye al desarrollo del MDM.



talend

Acciones de MDM



Acciones de MDM

- Divida su organización en *subject areas*.
- Priorizar las *subject areas*.
- Determinar la(s) fuente(s) de registro para cada *subject areas*.
- Determinar qué necesidad de *Governance* y en qué grado.
- Asignar valor a los distintos componentes del modelado de datos MDM:
 - Integración de datos, gestión de datos, calidad de los datos.



talend

Caso Práctico MDM



Caso Práctico MDM

Índice

- 1) Objetivo
- 2) MDM Open Studio
- 3) Crear Modelo de datos & Custom Types
- 4) Definir entidades
- 5) Create Views
- 6) Deploy
- 7) MDM Web User Interface
- 8) Implementar Jobs de sincronización
- 9) Definir triggers
- 10) Conclusiones

Objetivo



Objetivo

- Crear un repositorio centralizado definiendo 3 entidades.
- Cargar en repositorio datos procedentes de BBDD Vertica y PostgreSQL.
- Mantener actualizadas las BBDD Vertica y PostgreSQL ante cambios en el MDM.
- Proporcionar un formulario de introducción de datos.
- Mantener un registro de la evolución de los datos.



MDM Open Studio



MDM Open Studio - Introducción



- Unificar datos de clientes, productos, proveedores con una única *versión de la verdad*.
- La ruta más rápida hacia la gestión de datos maestros fiables y procesables.
- Como parte de *Talend Data Fabric*, MDM combina datos en tiempo real, aplicaciones e integración de procesos incluyendo reglas de calidad de datos.
- Administración para compartir los datos en aplicaciones locales, en la nube y en las aplicaciones móviles.
- Convertir datos en valor de negocio con una solución.

MDM Open Studio - Descarga

- Descarga: Talend Open Studio MDM, Master Data Management
 - <https://es.talend.com/products/mdm/mdm-open-studio/>
 - <https://www.talend.com/products/mdm/mdm-manuals-release-notes/>



MDM Open Studio - Instalación

- Una vez descargado el fichero, por ejemplo, *TOS_MDM-All-20181026_1147-V7.1.1.zip*
- Al descomprimirlo, hay otros dos ficheros:
- *TOS_MDM-Server-20181026_1147-V7.1.1.jar* : para instalar el Servidor.
 - Hay que instalarlo: en Windows con doble click (abrir con Java)
 - En Linux abrirlo con Java o ejecutar: `java -jar TOS_MDM-20181026_1147-V7.1.1.jar`
- *TOS_MDM-Studio-20181026_1147-V7.1.1.zip*: para instalar el *Studio* (únicamente hace falta descomprimir).



MDM Open Studio – Despliegue Server

- Linux:
 - `$/.../TOS_MDM-Server-20181026_1147-V7.1.1/apache-tomcat/bin$ sh startup.sh &`
 - `http://localhost:8180/talendmdm/ui`
- Windows:
 - `C:\TOS_MDM-Server-20181026_1147-V7.1.1/apache-tomcat/bin startup.bat`
 - `http://localhost:8180/talendmdm/ui`
- Usuario (defecto): user/user
- Admin (defecto): administrator/administrator



MDM Open Studio – Consola server

The screenshot displays the Talend MDM Community Edition interface. The top navigation bar includes the Talend logo, the text 'TALEND MDM COMMUNITY EDITION', and user information: 'administrator', 'Español', and a 'Logout' button. The main content area is titled 'Master Data Browser' and shows a table of product data. The table has columns for 'Id', 'Picture', 'Name', 'Description', and 'Features'. The data is filtered by 'Product' and 'Id' with the condition 'is equal to *'. The table lists 20 products, including Wiza LLC, Bartoletti LLC, Reilly, Cruickshan..., Collins Inc, Ondricka Inc, Cartwright Inc, Okuneva-Schawal..., Botsford, Haley an..., Hahn-Moen, Morar-Lesch, Langworth, Howel..., Mraz, Schulist and..., Senger-Huel, Raynor Group, Fahey LLC, Roberts LLC, Wilderman-Bartol..., Mayer LLC, Leuschke Group, and Boyle Group. On the right side, there are two configuration panels: 'Domain Configuration' and 'Portal Configuration'. The 'Domain Configuration' panel shows 'Data Container: Producto' and 'Data Model: Producto' with a 'Save' button. The 'Portal Configuration' panel shows 'Widgets:' with 'Getting Started' and 'Processes' (checked) and a 'Save' button. The bottom status bar indicates 'Page 1 of 50' and 'Lines/page: 20'.

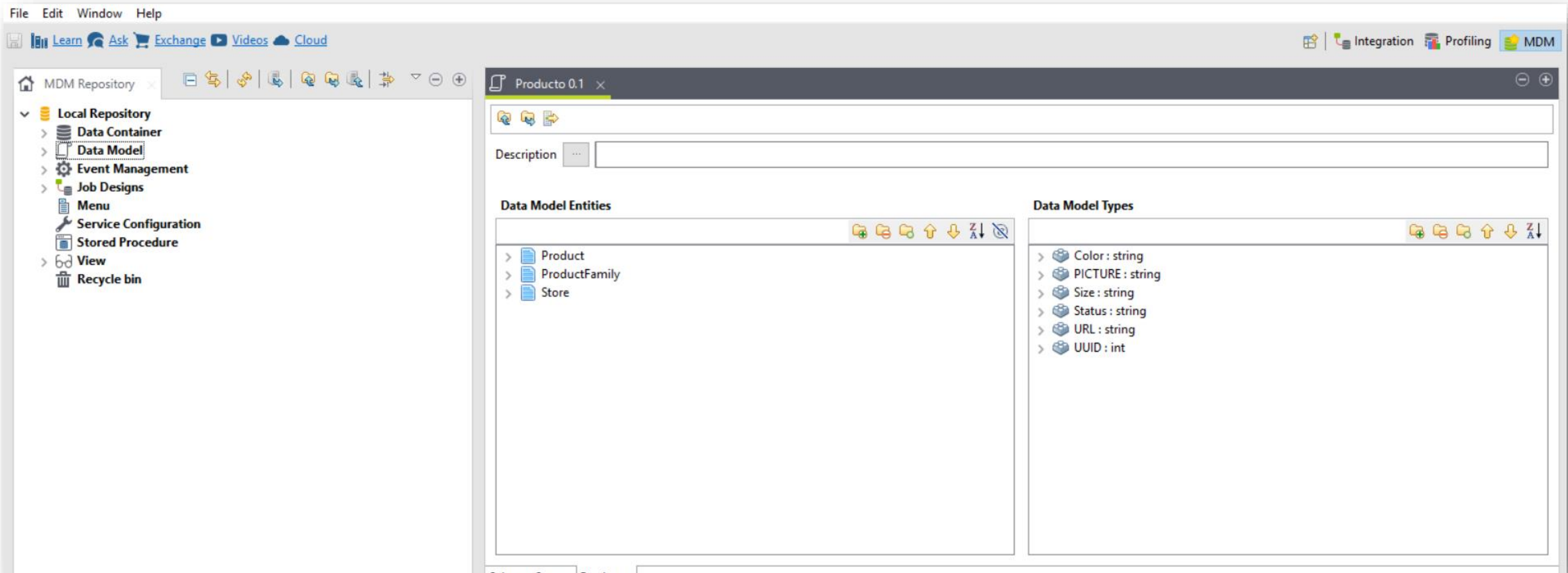
Id	Picture	Name	Description	Features
1		Wiza LLC	Elantra	Hill Manufacturing...
2		Bartoletti LLC	C8 Spyder	CVS Pharmacy
3		Reilly, Cruickshan...	CX-7	Physicians Total C...
4		Collins Inc	Savana 1500	NCS HealthCare of...
5		Ondricka Inc	Eos	Cardinal Health
6		Cartwright Inc	1500	Lupin Pharmaceut...
7		Okuneva-Schawal...	Talon	Amar Remedies Li...
8		Botsford, Haley an...	Diamante	Bare Escentuals B...
9		Hahn-Moen	Escalade ESV	Advanced Beauty ...
10		Morar-Lesch	Tredia	Revamp Internati...
11		Langworth, Howel...	Ram Van 2500	St Marys Medical ...
12		Mraz, Schulist and...	Q7	Caraco Pharmace...
13		Senger-Huel	Wrangler	McKesson (Sunma...
14		Raynor Group	944	Mylan Pharmaceu...
15		Fahey LLC	Grand Am	Mylan Pharmaceu...
16		Roberts LLC	Type 2	Kroger
17		Wilderman-Bartol...	164	St Marys Medical ...
18		Mayer LLC	S60	Nelco Laboratorie...
19		Leuschke Group	Safari	Ranbaxy Pharmac...
20		Boyle Group	600	Dispensing Soluti...

MDM Open Studio – Despliegue Studio

- Haga doble clic en el archivo ejecutable correspondiente a su sistema operativo, por ejemplo:
- TOS_* - win-x86_64.exe, para Windows.
- TOS_* - linux-gtk-x86_64, para Linux.
- TOS_* - macosx-cocoa.app, para Mac.

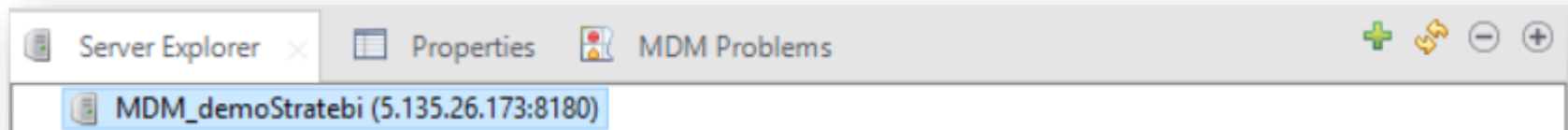


MDM Open Studio – Despliegue Studio



MDM Open Studio – Conexión a MDM

- El primer paso es crear una conexión al servidor MDM.
 - En la ventana principal de Studio, en el panel *Server Explorer*, haga clic en el botón + para especificar un nuevo servidor MDM



Add Server Location

Name: My Server

Server: http://localhost:8180/talendmdm/services/soap

Authentication

Username: administrator

Password:

☒ Share Password

Check Connection OK Cancel



MDM Open Studio – Gestión de datos

- Los siguientes pasos serían de *Data Governance Tasks* (Tareas de gestión de datos).
- Validar los datos del origen contra el modelo de datos antes de guardarlos como datos maestros, para asegurarnos que están limpios, precisos y coherentes.
- Vamos a utilizar un ejemplo de datos llamado *Products*, con tres entidades:
 - *Products*
 - *ProductFamily*
 - *Store*
- Cada uno de ellos con sus elementos o entidades.

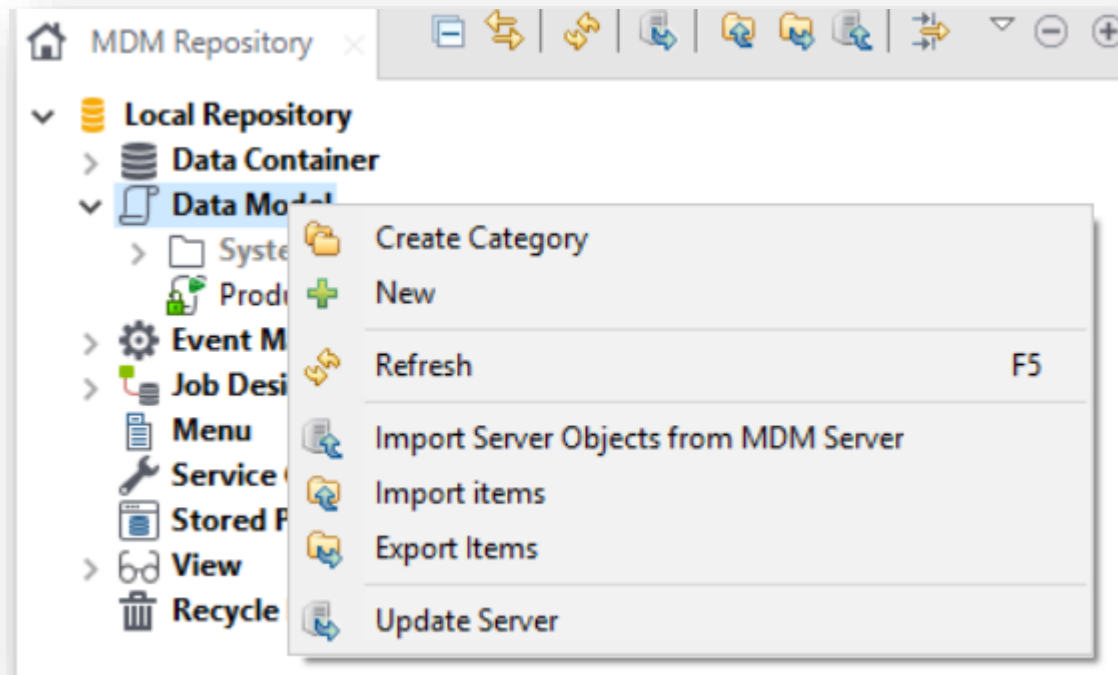


Crear Modelo de datos & Custom Types



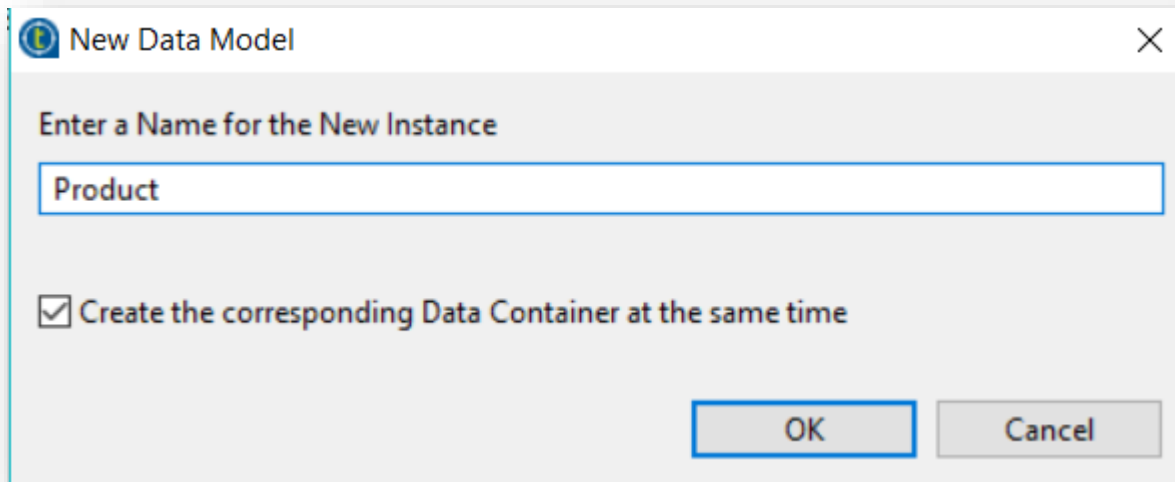
Crear Modelo de datos & Custom Types

- Crear el Modelo de Datos (*Data Model*).



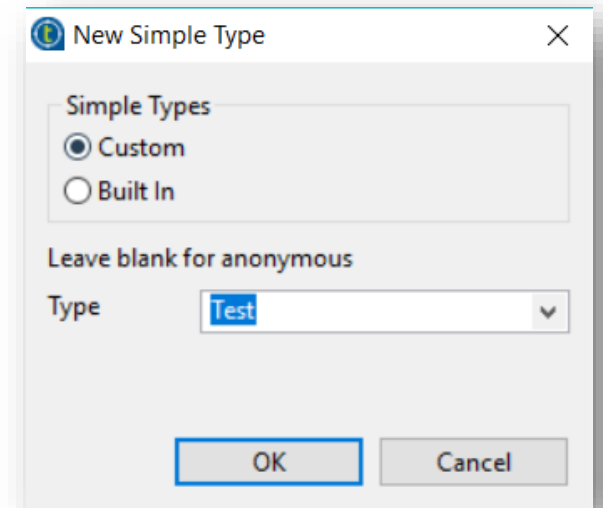
Crear Modelo de datos & Custom Types

- Introducir un nombre para el modelo, por ejemplo, Producto:
 - Seleccionar la opción de *Create the corresponding Data Container at the same time*.



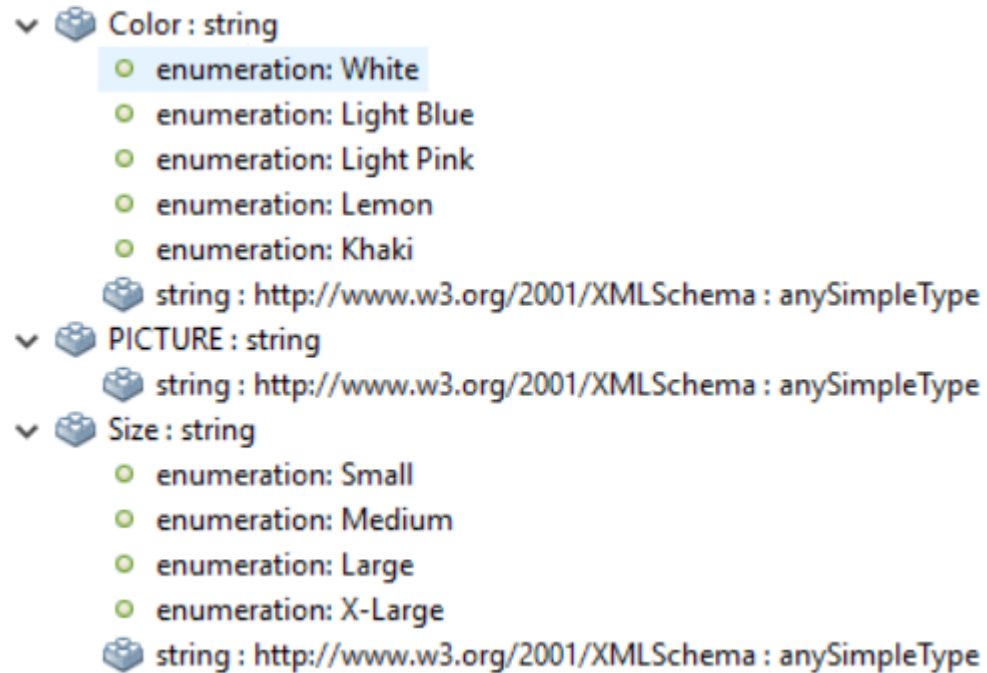
Crear Modelo de datos & Custom Types

- Para definir atributos de tipos personalizados, puede cambiar un atributo de tipo simple a un valor personalizado predeterminado
- Escriba o defina un nuevo tipo personalizado y utilícelo para los atributos.
- Definimos los siguientes *Custom Types*:
 - Color
 - PICTURE
 - Size
 - Status
 - URL

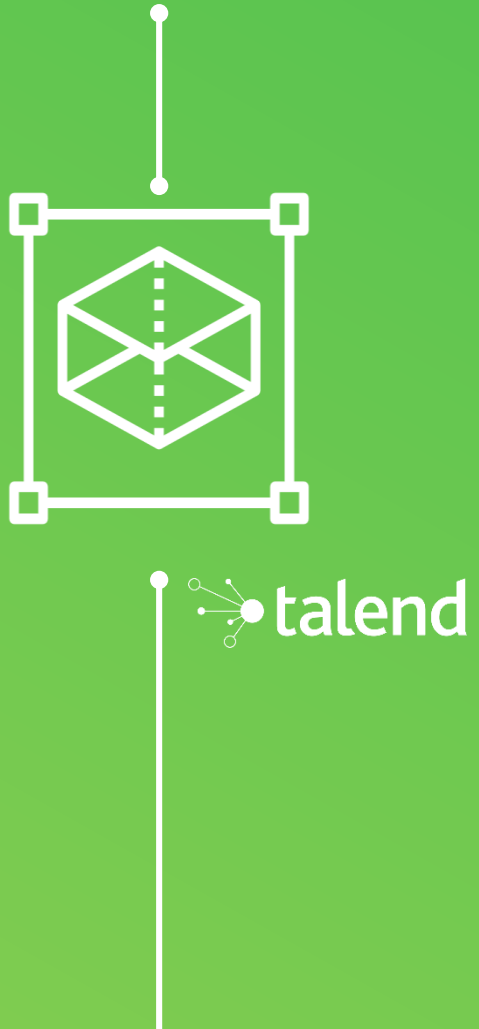


Crear Modelo de datos & Custom Types

- Para definir *enumeration*, *right click* sobre el *Type* en concreto y *Edit Enumeration*.

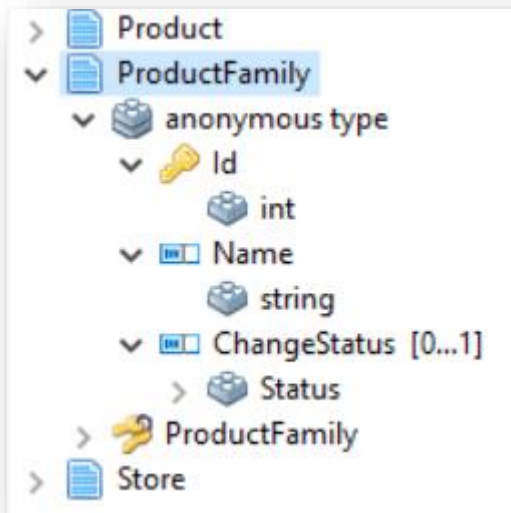


Definir Entidad ProductFamily



Definir Entidad ProductFamily

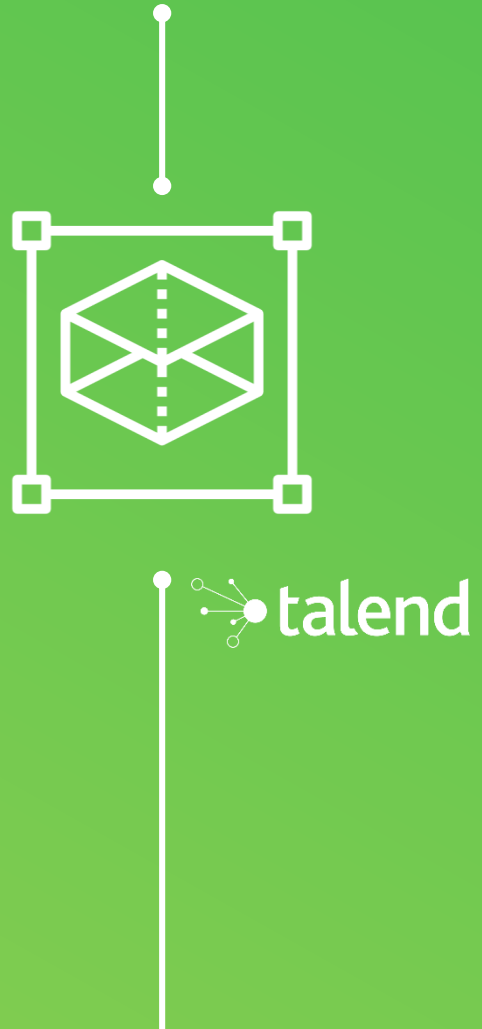
- Crear *New Entity: ProductFamily*:



Element Name	Element Type	Min. Ocurrance	Max. Ocurrance
Id (auto)	int	1	1
Name	String	1	1
ChangeStatus	String	0	1

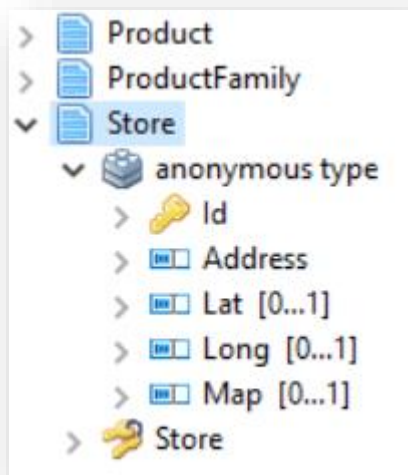


Definir Entidad Store



Definir Entidad Store

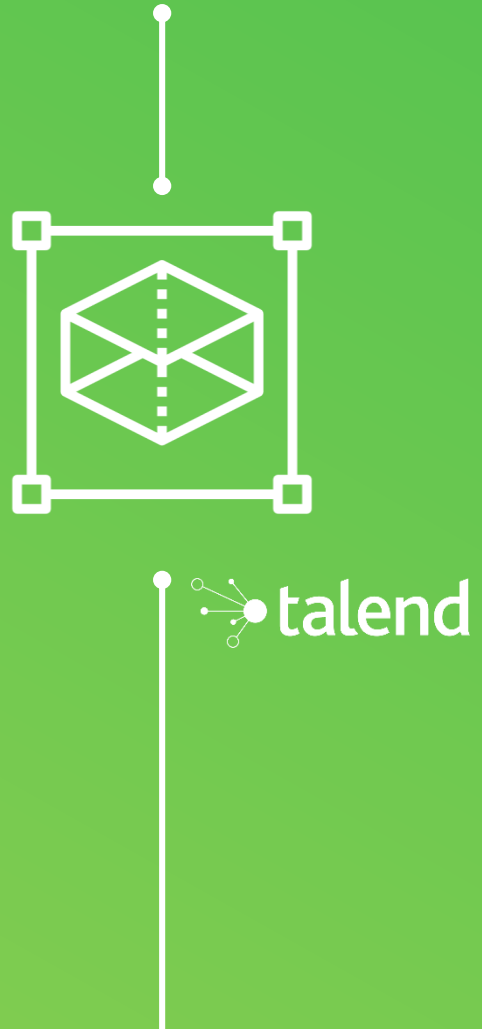
- Crear *New Entity: Store*:



Element Name	Element Type	Min. Ocurrance	Max. Ocurrance
Id (auto)	int	1	1
Address	String	1	1
Lat	Double	0	1
Long	Double	0	1
Map	URL	0	1

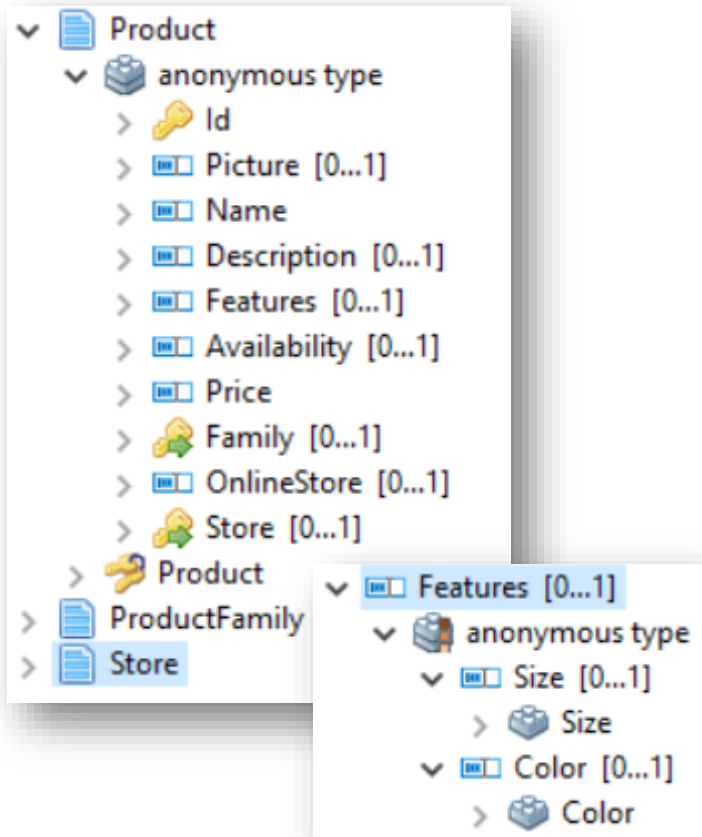


Definir Entidad Product



Ejercicio 3. Crear Entidad Product

- Crear *New Entity: Product*:



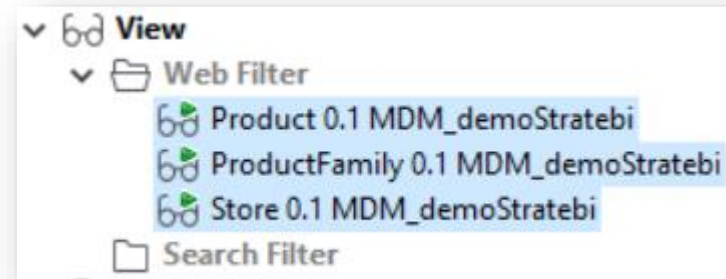
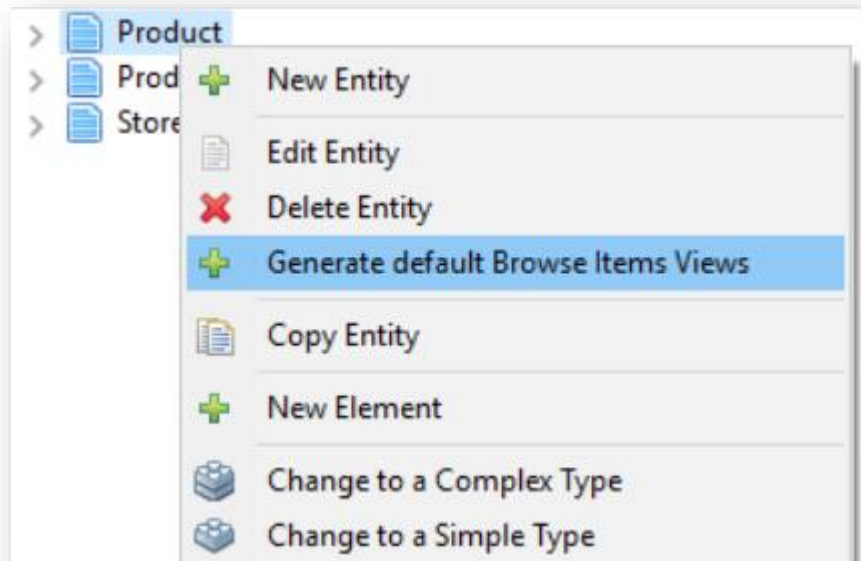
Element Name	Element Type	Min. Occurrence	Max. Occurrence
Id (auto)	int	1	1
Picture	String	0	1
Name	String	1	1
Description	String	0	1
Features	Anonymous type	0	1
Features/Size	Size	0	1
Features/Color	Color	0	1
Availability	boolean	0	1
Price	Decimal	1	1
Family (FK)	String	0	1
OnlineStore	URL	0	1
Store (FK)	String	0	1

Create Views



Create Views

- Crear vistas para:
 - Product
 - ProductFamily
 - Store



talend

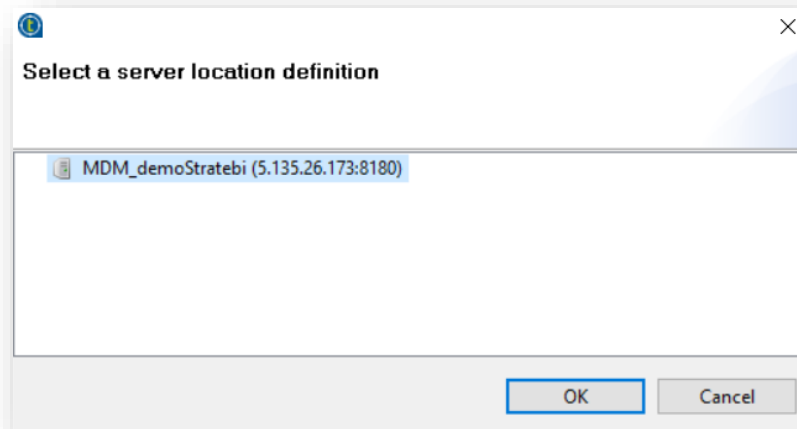
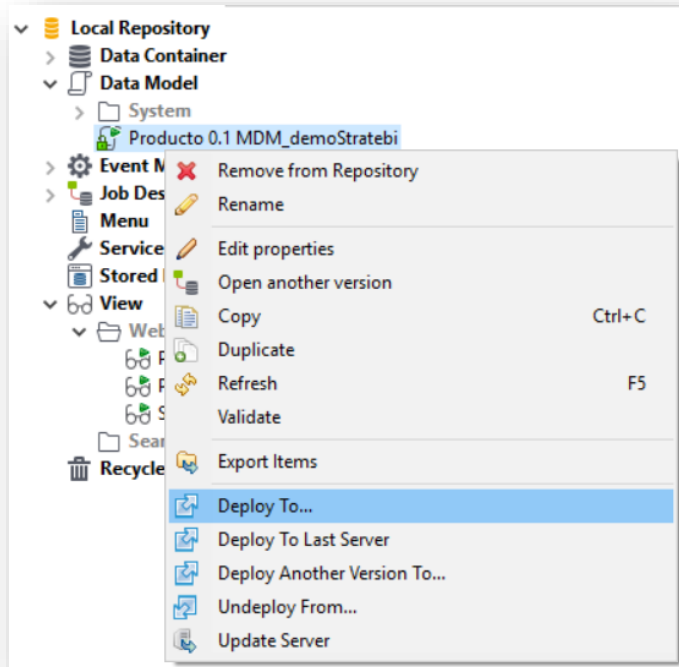
Deploy



talend

Deploy

- **Siempre** hay que hacer un *deploy* de los objetos MDM: data model, data container y views
 - Si no se hace *deploy* el servidor no 'sabe' de ellas en ejecución.
 - Pasos: selecciona el objeto, right click, 'Deploy To....'



MDM Web User Interface



MDM Web User Interface

- Acceso sobre <http://localhost:8180/talendmdm/ui>
- Login credenciales: **administrator/administrator**



Login

Password



MDM Web User Interface

- Sobre *Actions* en la parte de la derecha, se obtiene *Domain Configuration*:
 - Seleccionar el *Data Container* subido y seleccionar el modelo asignado.
 - Finalmente pulsar en *Save*.
- En *Master Data Browser*, se podrán ver los datos disponibles, añadir nuevos, modificar, actualizar o eliminar.

Actions >>

Domain Configuration

Data Container:

Producto ▼

Data Model:

Producto ▼

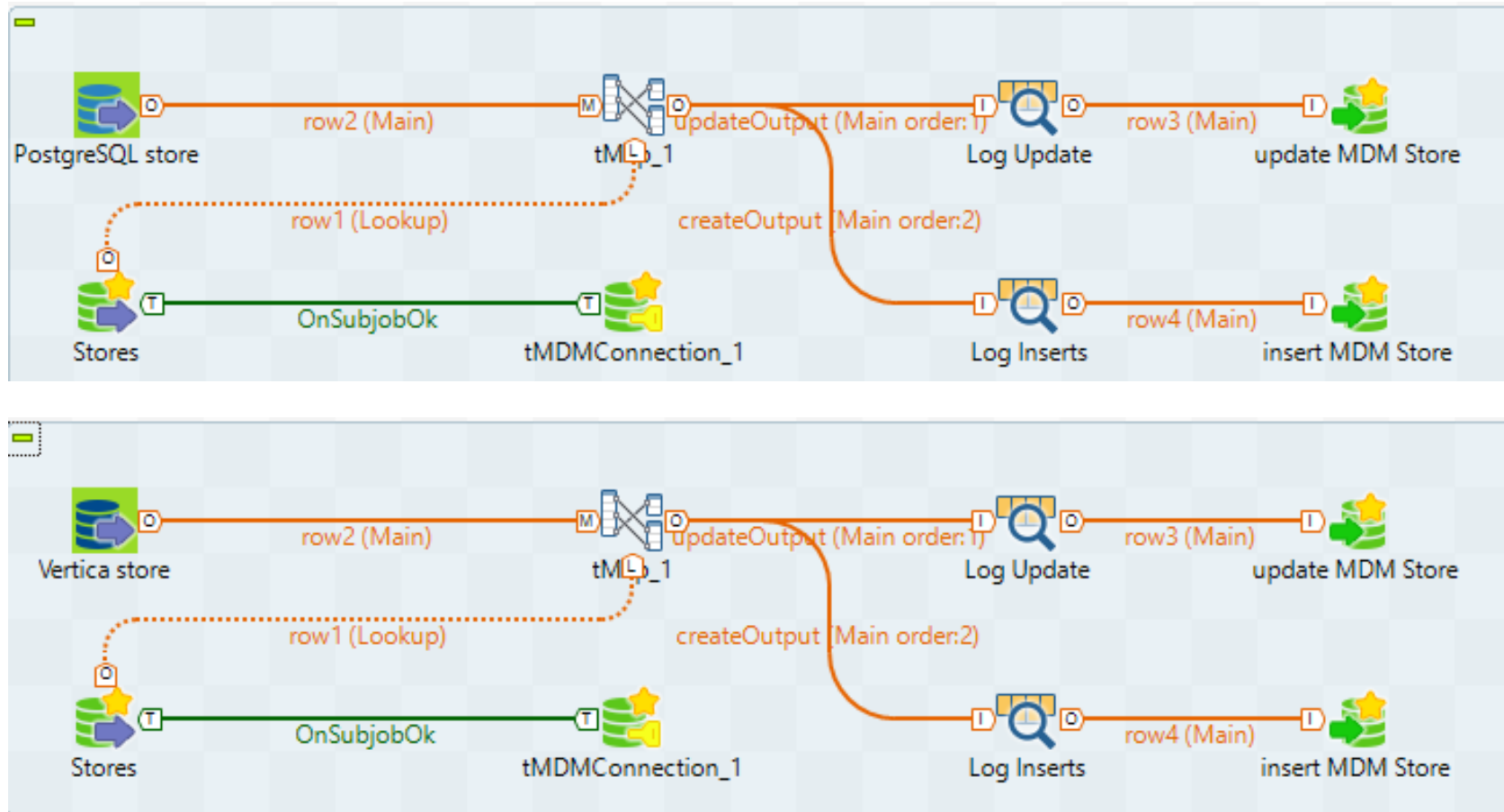
Save

Master Data Browser ✕					
+ Create ✕ Delete ▾ 📄 Import and Export ▾ Product ▼ Id ▼ is equal to ▼					
☐	Id	Picture	Name	Description	Features
☐	1	Wiza LLC	Elantra	Hill Manufacturing...	
☐	2	Bartoletti LLC	C8 Spyder	CVS Pharmacy	
☐	3	Reilly, Cruickshan...	CX-7	Physicians Total C...	
☐	4	Collins Inc	Savana 1500	NCS HealthCare of...	
☐	5	Ondricka Inc	Eos	Cardinal Health	
☐	6	Cartwright Inc	1500	Lupin Pharmaceut...	
☐	7	Okuneva-Schowal...	Talon	Amar Remedies Li...	
☐	8	Botsford, Haley an...	Diamante	Bare Escentuals B...	

Job – Carga de Store Entity



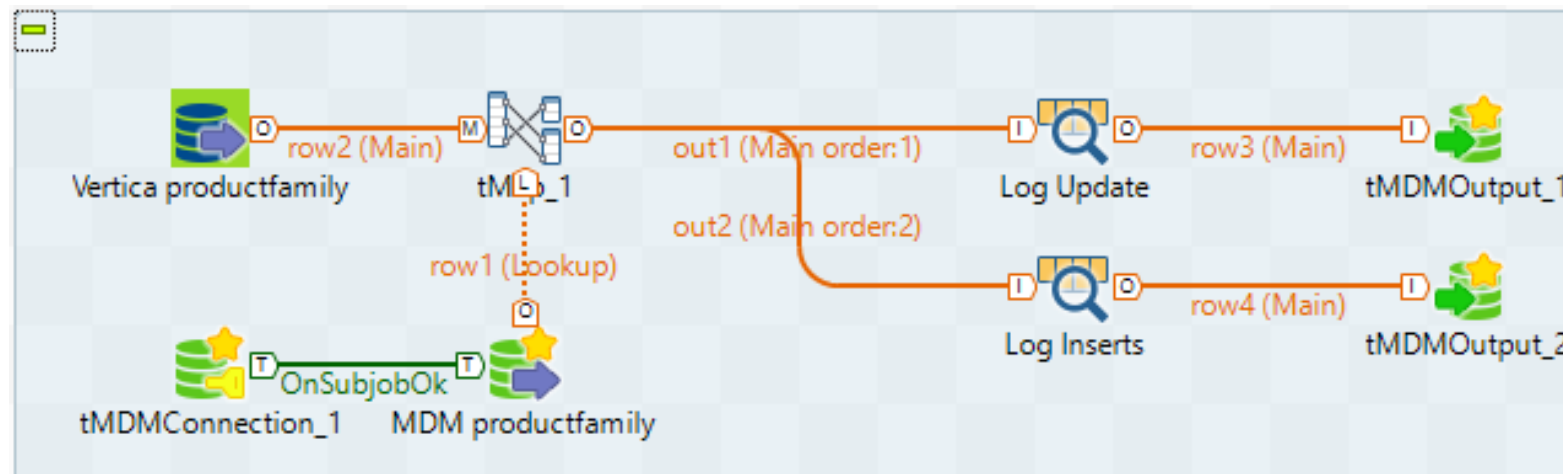
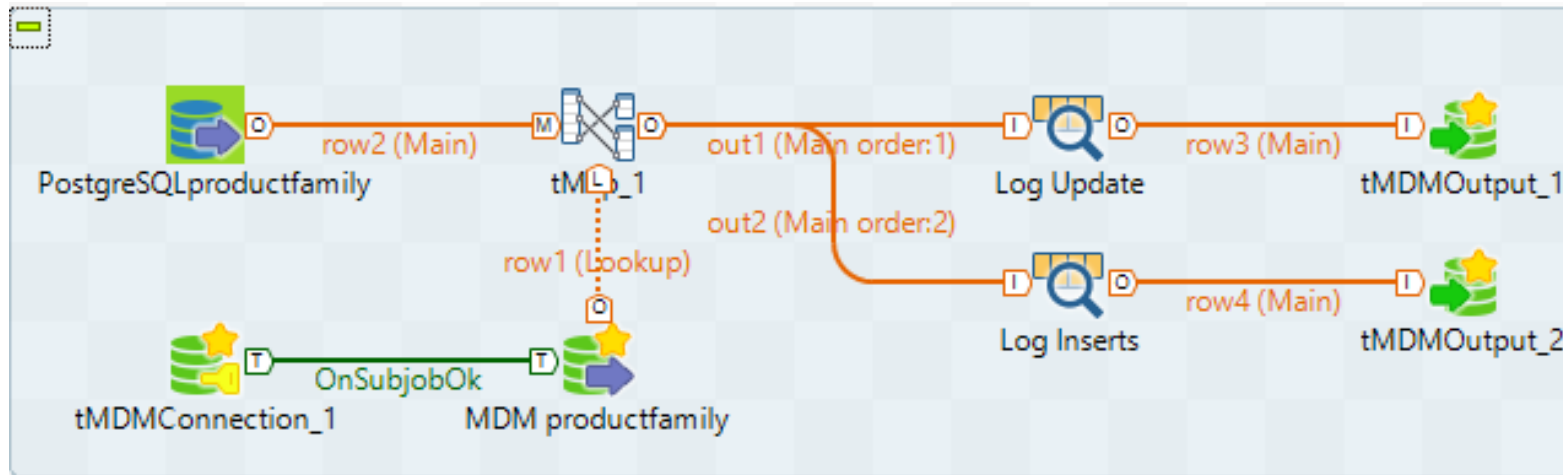
Jobs – Carga de Store Entity en MDM



Job – Carga de ProductFamily Entity



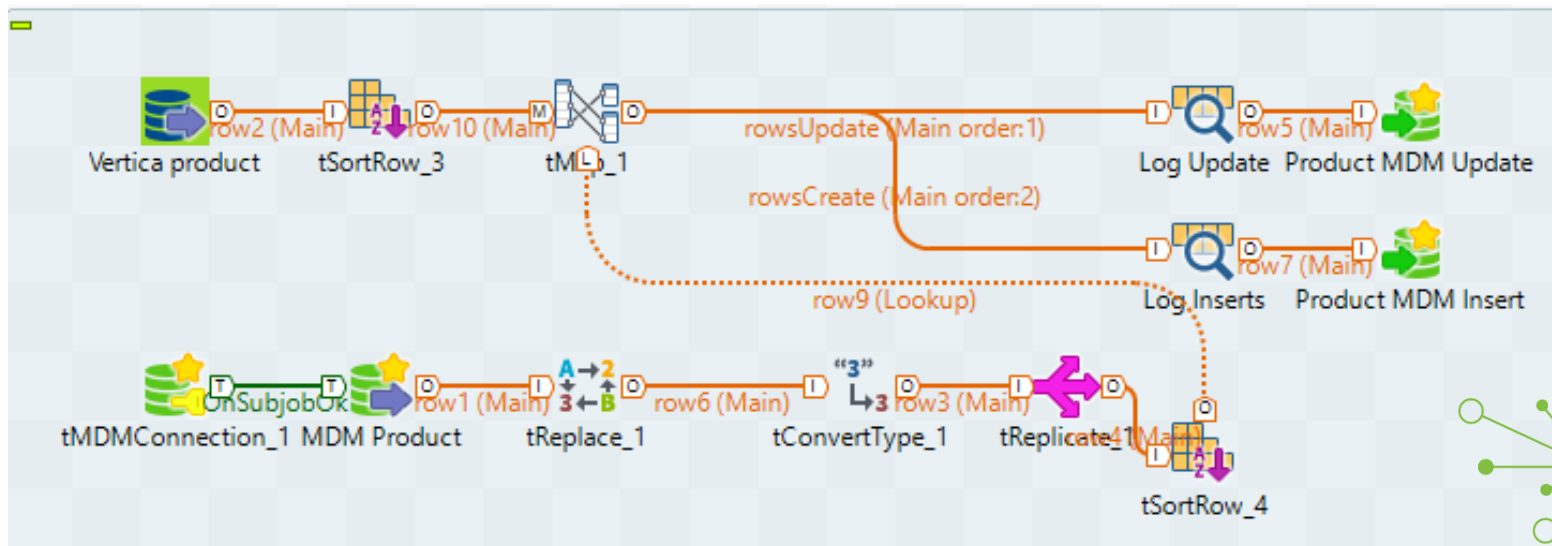
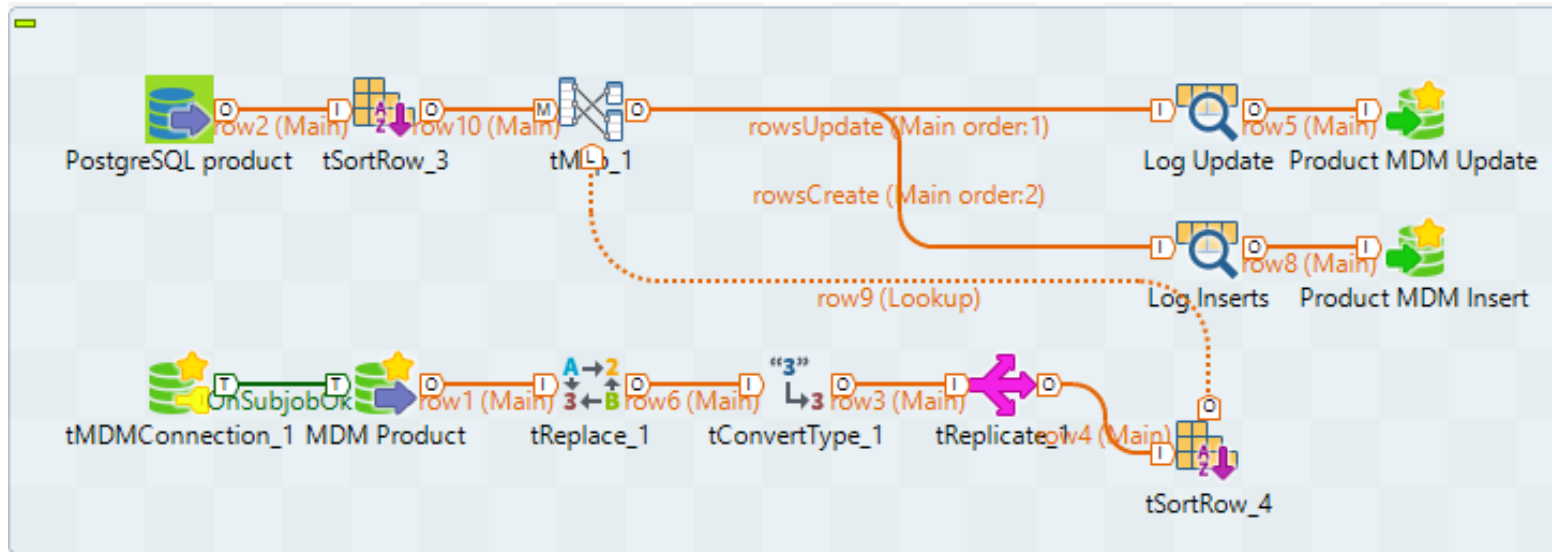
Jobs – Carga de ProductFamily Entity



Job – Carga de Product Entity



Jobs – Carga de Product Entity



Triggers MDM



Triggers MDM

- Como ya se definió en la arquitectura de MDM, hay que mantener la integridad de datos en cada uno de los orígenes de datos conectados al servidor.
- Para alcanzar este objetivo, se disponen de distintos acercamientos. En las versiones mas recientes de Talend MDM se introdujo un sistema de publicación de triggers mediante los componentes:
 - tMDMTriggerInput u tMDMTriggerOutput: Se interpreta el tipo de operación y se actúa en consecuencia



Triggers MDM

- Las acciones se pueden seguir desde la interfaz web quedando registradas en el *journal* (diario). Para mantener la integridad de las bases de datos se pueden tratar las operaciones *CREATE*, *UPDATE* & *PHYSICAL_DELETE*. En consecuencia hay que desarrollar los Jobs necesarios para el tratamiento de cada *entidad*.

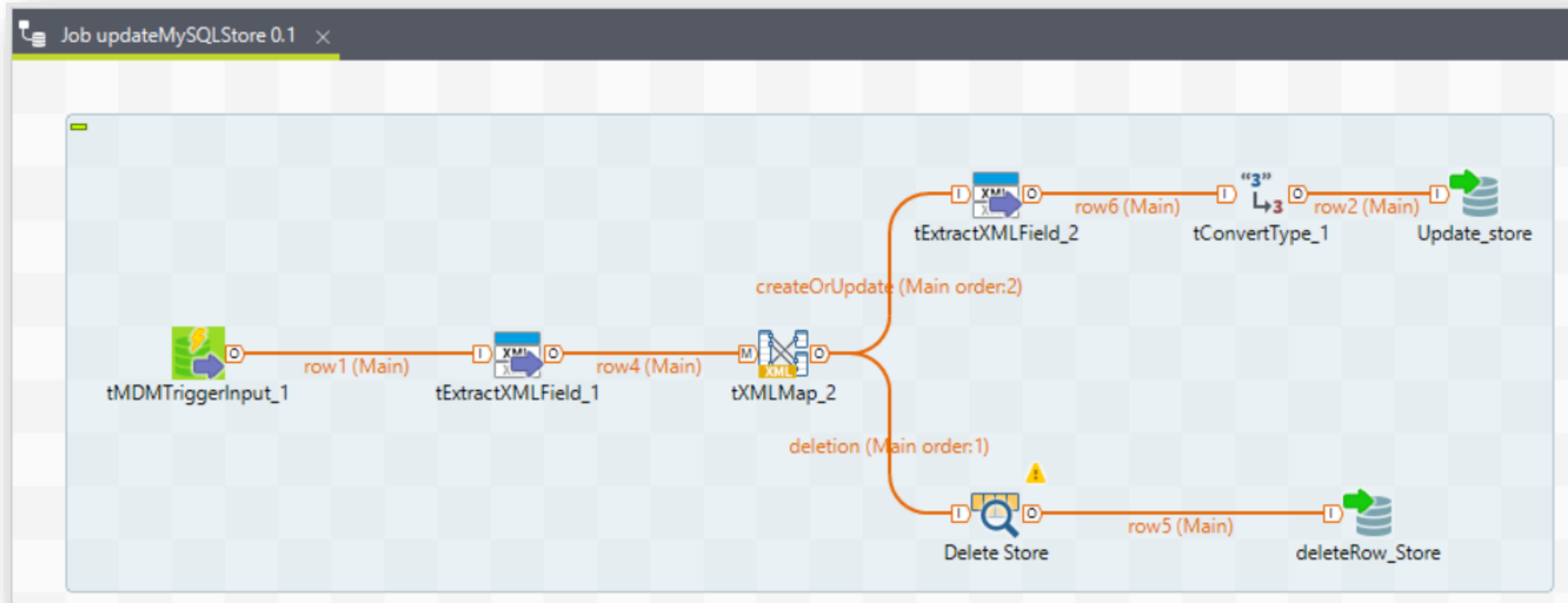
The screenshot shows the 'Journal' window in Talend MDM. It has a 'Search Panel' with filters for Data Model (set to 'Producto'), Entity, Start Date, and End Date. On the right, there are fields for Key, Operation Type (with a dropdown menu open showing options: ALL, CREATE, UPDATE, PHYSICAL_DELETE, LOGIC_DELETE, RESTORED, ACTION), and Source. A checkbox for 'Strict S' is checked. At the bottom, there are tabs for 'Results' and 'Timeline', and a table header with columns: Data Container, Data Model, Entity, Key, Operation Type..., Operation Time, Source, and Us.

Trigger - Store



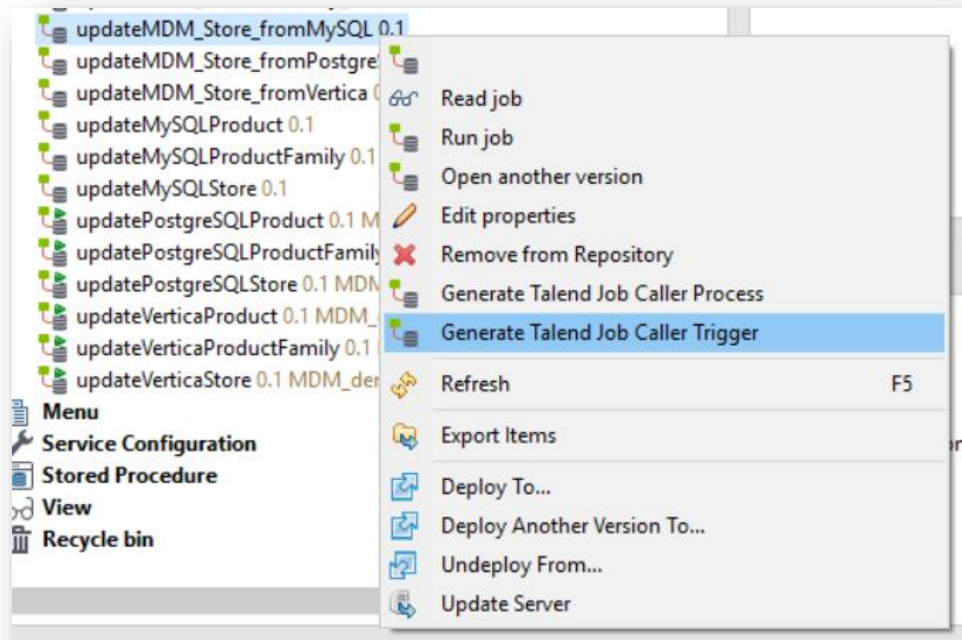
Trigger - Store

- Jobs (uno para PostgreSQL y otro para Vertica, cambiar únicamente los pasos de salida):



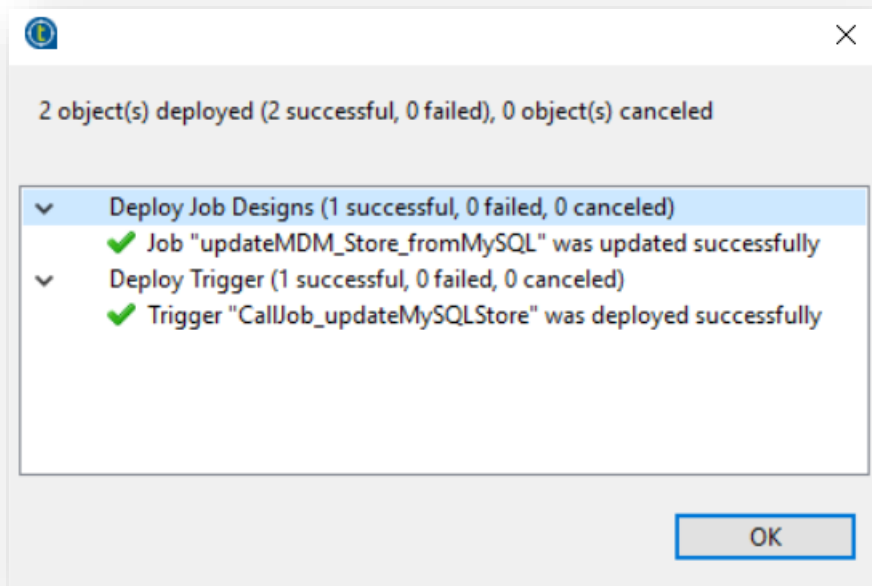
Trigger Deploy - Store

- Una vez, finalizado el desarrollo del Job, hay que realizar un *deploy* del Job en el servidor y de su correspondiente *trigger*. Para ello, se cambiará a la perspectiva MDM y se creará el *trigger* pertinente realizando un *click* derecho sobre el Job requerido.



Trigger Deploy - Store

- Finalmente, realizar un *deploy* tanto del *trigger* generado como del *Job* al servidor MDM. Y probar a realizar una modificación sobre la Entidad Store.



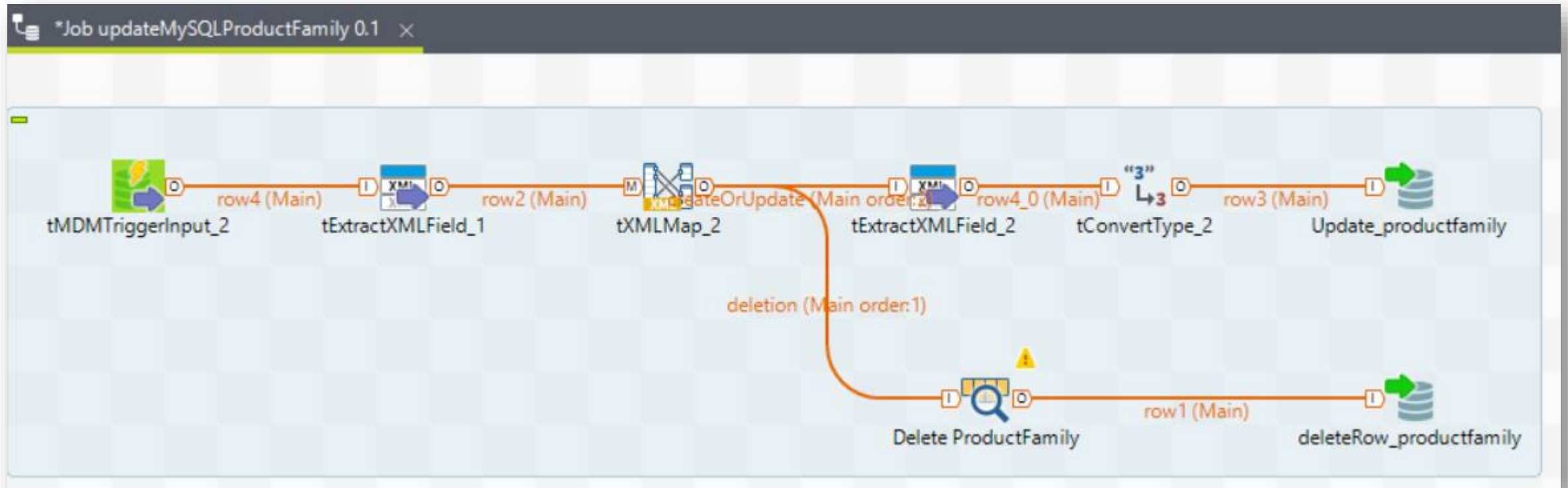
```
15:07:10,546 INFO [JobDeploy] Job updateMDM_Store_fromMySQL_0.1.zip has been successfully deployed.  
15:07:10,550 INFO [JobContainer] Adding new class loader com.amalto.core.jobbox.util.JobClassLoader@c0964ca
```

Trigger – ProductFamily



Trigger - ProductFamily

- Jobs (uno para PostgreSQL y otro para Vertica, cambiar únicamente los pasos de salida):

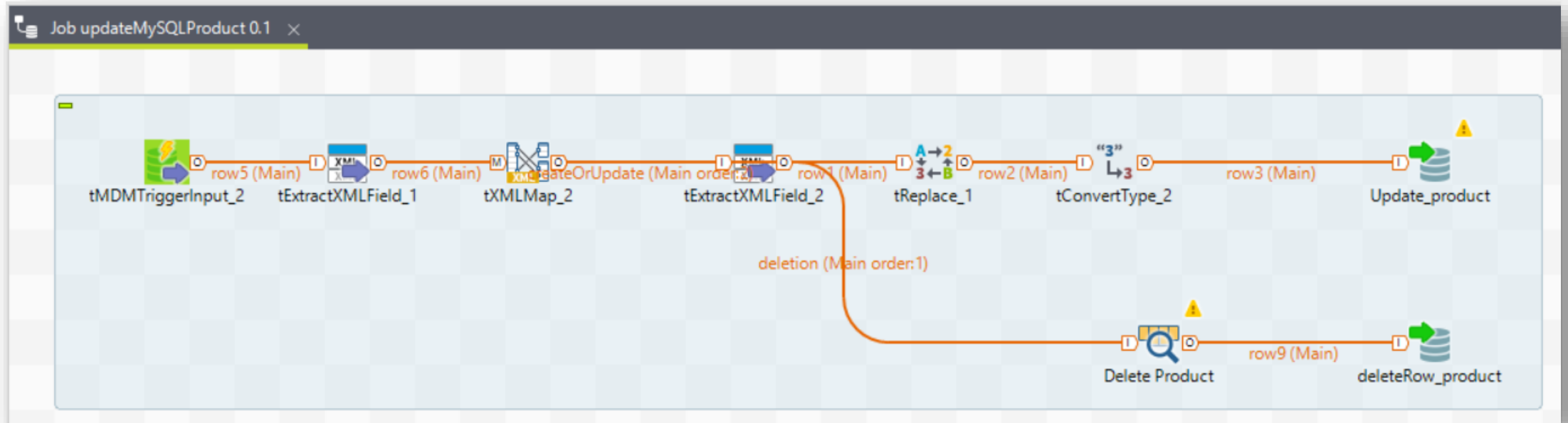


Trigger – Product



Trigger - Product

- Jobs (uno para PostgreSQL y otro para Vertica, cambiar únicamente los pasos de salida):



Conclusiones



Conclusiones

- Fácil trazabilidad del dato mediante consola y/o logs.
- Sincronización directa desde MDM a las BBDD mediante triggers.
- ¿Sincronización desde las BBDD a MDM?.



Preguntas

