



Gobernanza y gestión del dato con TALEND

WEBINAR

Contenido Teórico

Índice

1) DATA HEALTH

2) GOBERNANZA DEL DATO

3) DATA MANAGEMENT

4) TALEND DATA CATALOG

5) TALEND DATA PREPARATION & TALEND DATA STEWARDSHIP



DATA HEALTH



DATA HEALTH

And data remains a critical problem

78%

of executives have challenges making data-driven decisions

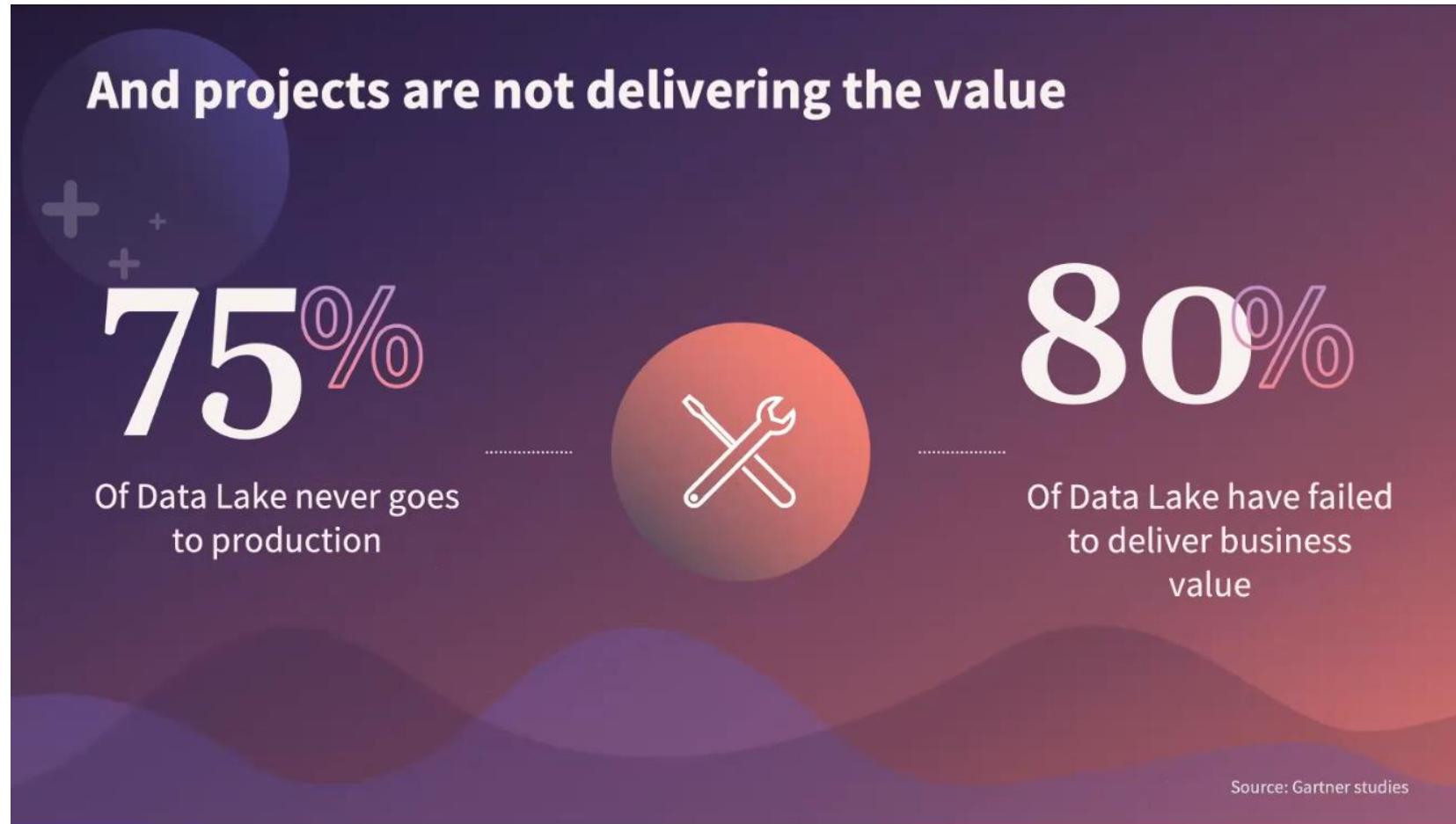


60%

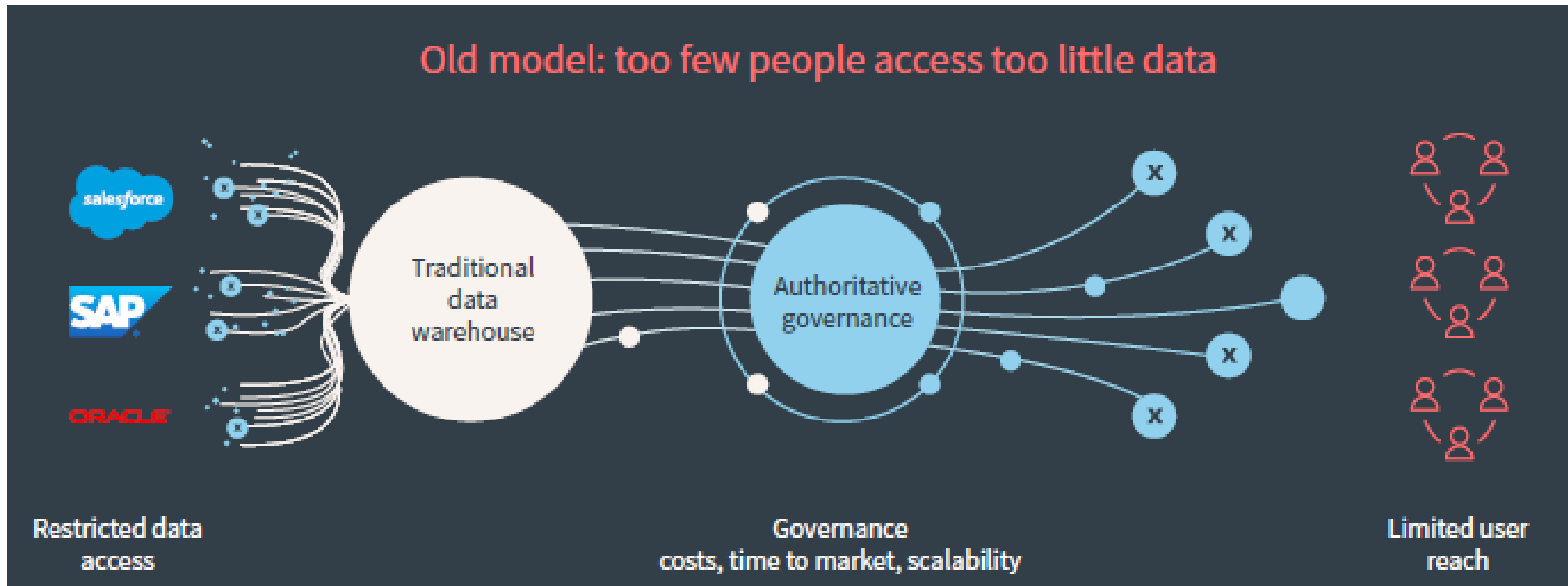
of executives **DON'T** always trust the data they work with

Source: Talend 2021 Data Health Survey

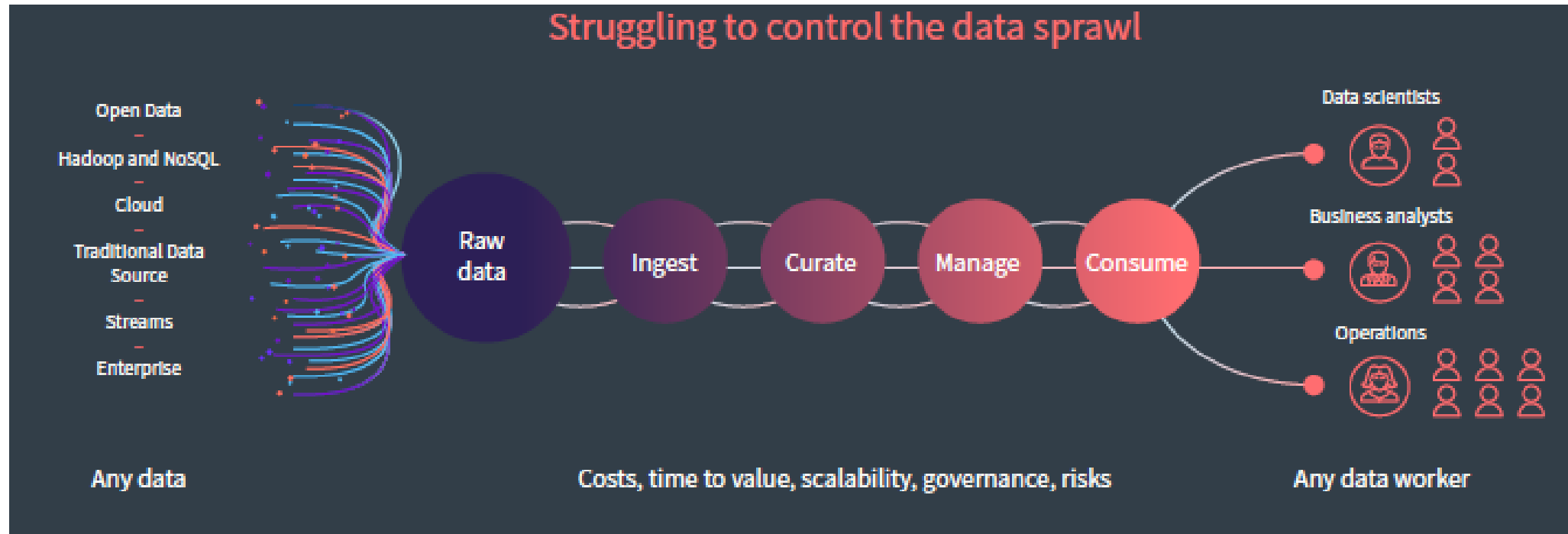
DATA HEALTH



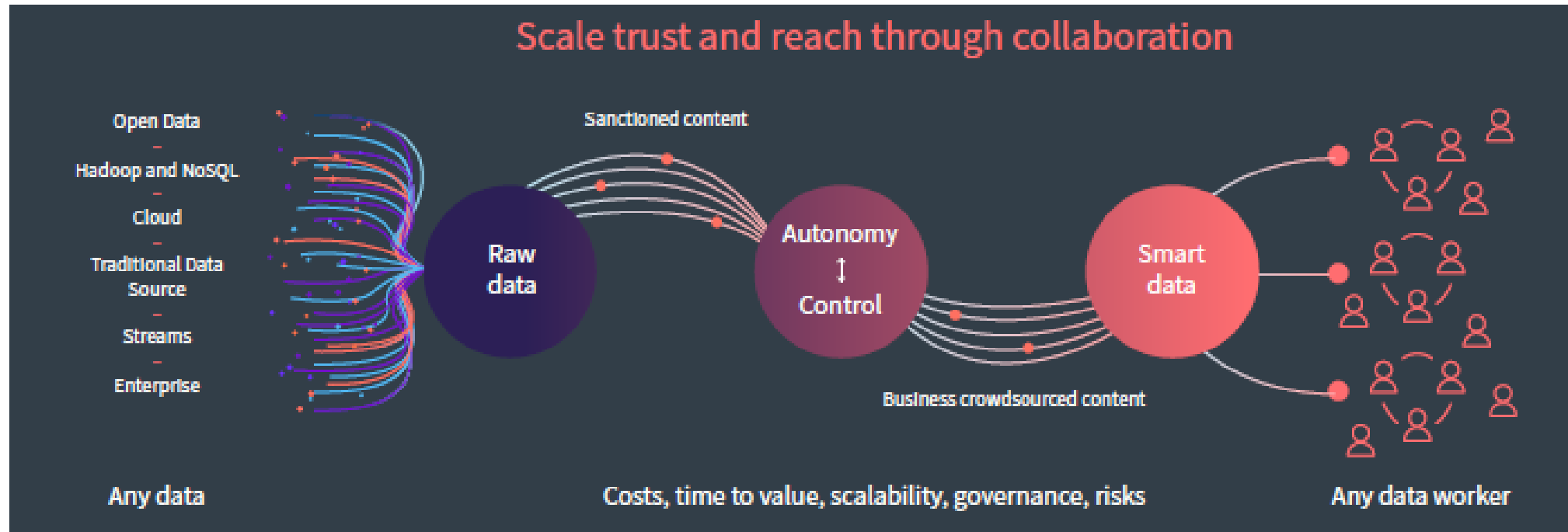
DATA HEALTH



DATA HEALTH



DATA HEALTH



DATA HEALTH

Data health is our vision for the enterprise



Data health:

Data health describes how well a company's data supports its business objectives



Preventative Measures

Preemptively identify and resolve data challenges



Effective Treatments

Systematically improve data reliability and reduce risk



Supportive Culture

Establish an organizational discipline around data care and maintenance



Proactive monitoring and reporting

DATA HEALTH

A Maturity Model for Data Health

Data-driven

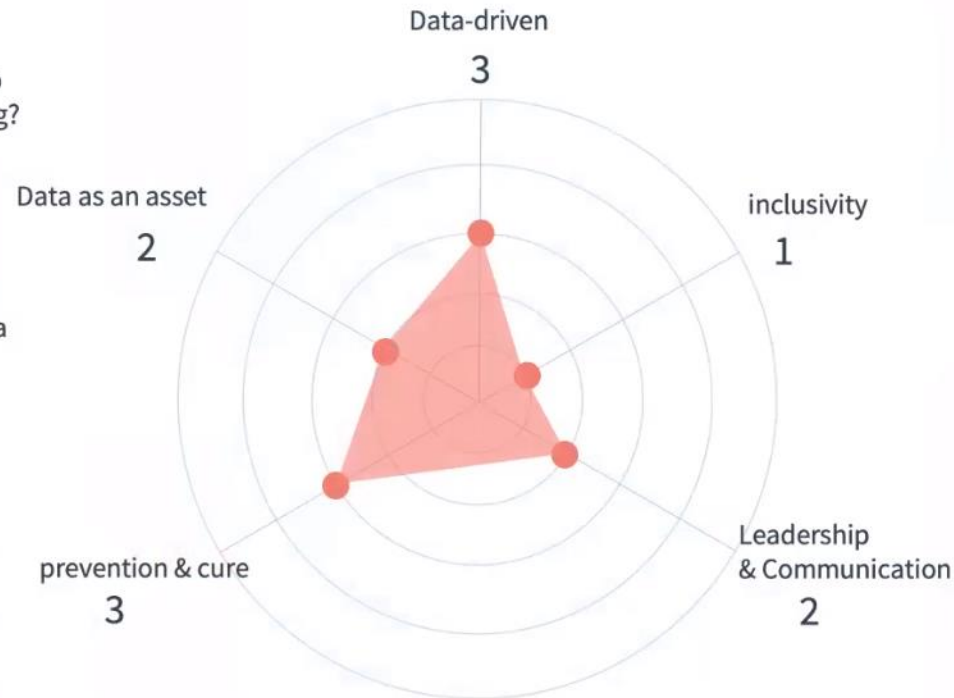
How well is data used to support decision making?

Data as an Asset

How is data treated as a corporate asset?

Prevention & Cure

How well does the organization speak the language of data?



Inclusivity

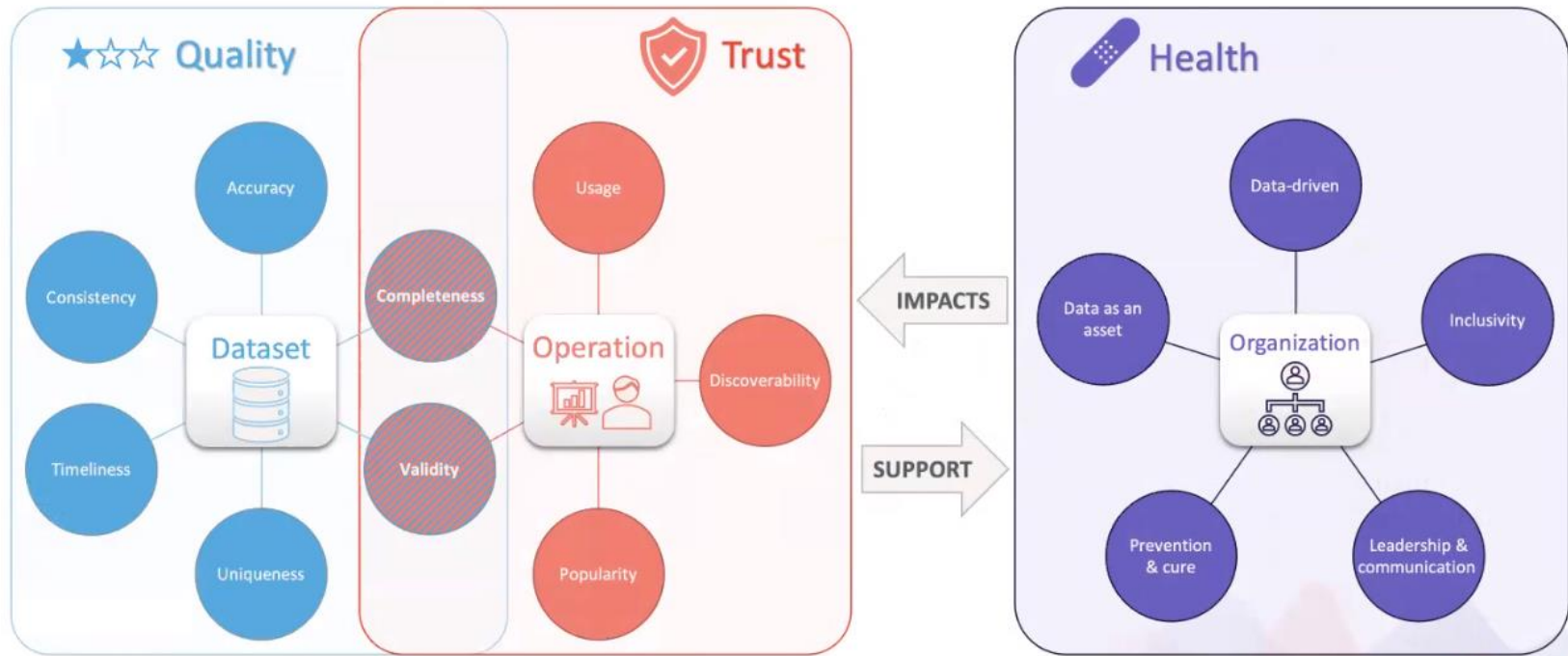
Are all employees able to do the right thing for data health?

Leadership & Communication

How well is data health strategy and execution communicated?
How well do leadership support data health?

DATA HEALTH

Data Health compare to Data Quality & Data Trust



DATA HEALTH

Delivering Actionable Insight

Your path to Data Health is unique to you



GOVERNANZA DEL DATO



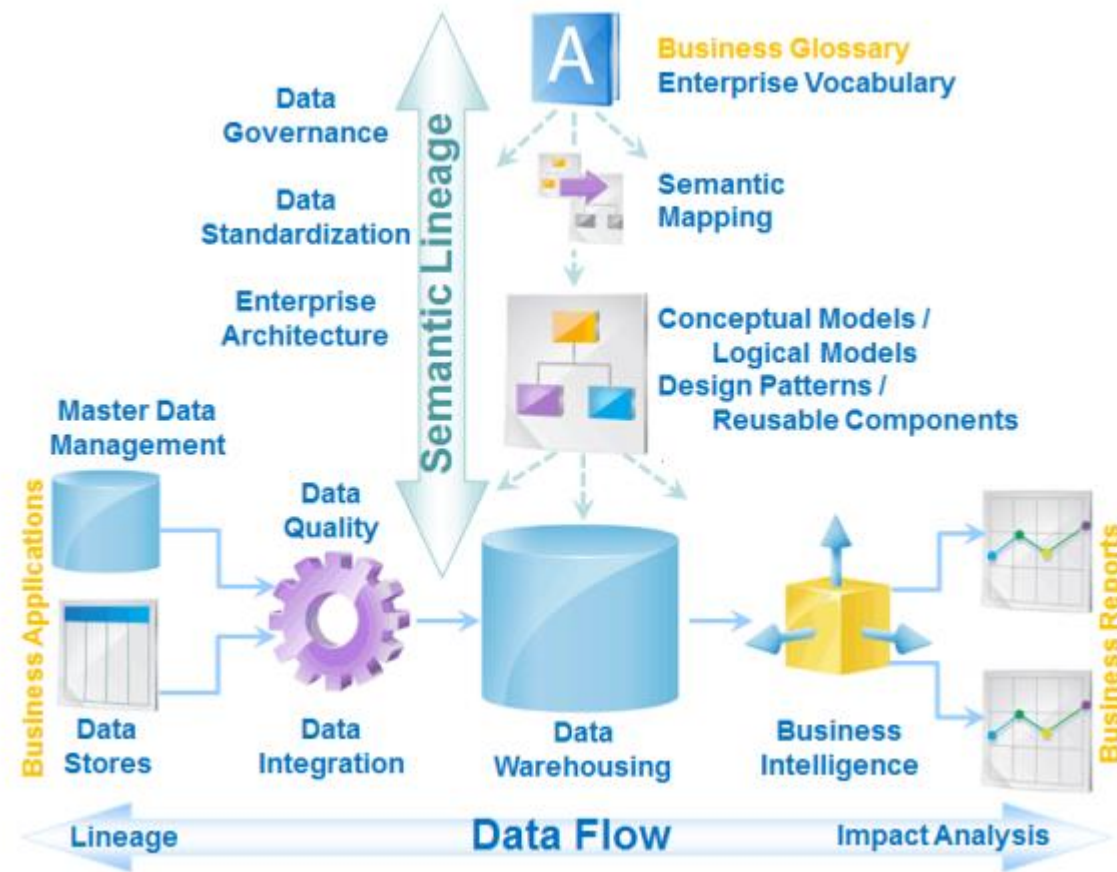
GOBERNANZA DEL DATO

- La gobernanza del dato en el BI es el uso eficaz de la información aportada por los datos, a partir del cumplimiento de una serie de procesos, normas, funciones y políticas.
- Una organización utiliza la gobernanza del dato para garantizar que los datos con los que trabaja, logren llevarlos al éxito mediante la toma de decisiones seguras y de confianza.



TALEND STRUCTURE

La gobernanza del dato es un proceso dentro del Data Management, ya que se ocupa de la planificación, supervisión y control en la gestión y uso de datos.



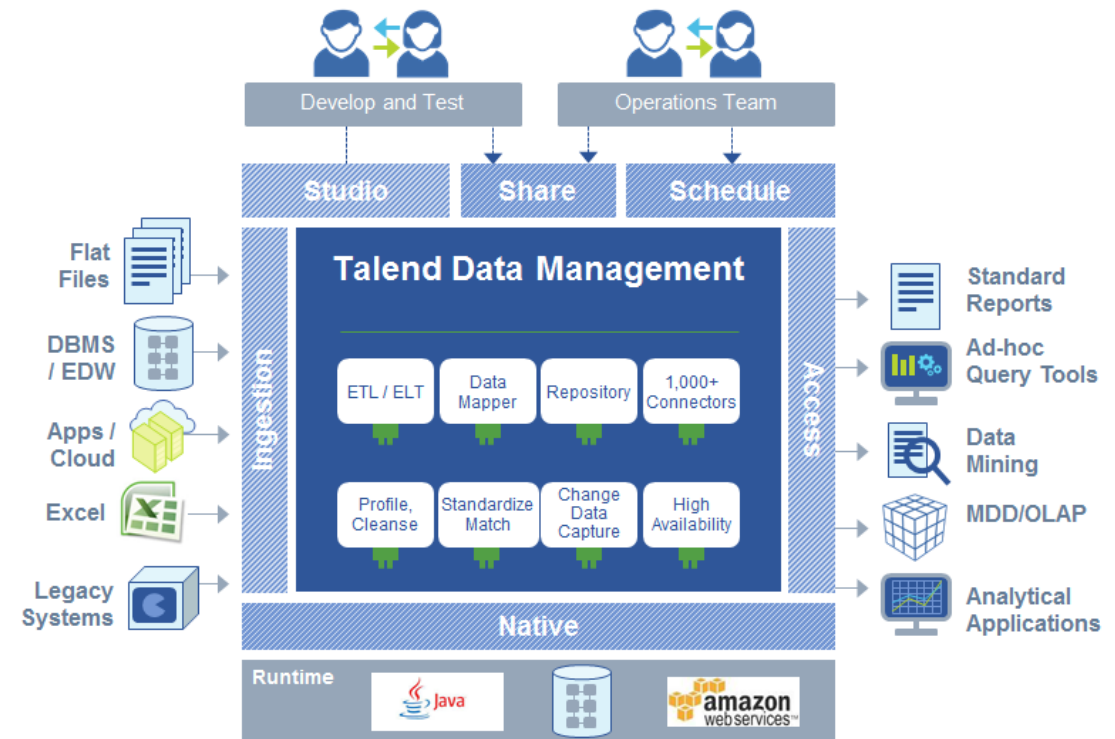
DATA MANAGEMENT



DATA MANAGEMENT

El *data management* o *MDM* se caracteriza por la unión de todos los datos críticos de una organización en un solo repositorio maestro.

En un nivel más alto, esta descripción se corresponde con una gestión flexible y bien gestionada de los datos de una organización, garantizando la mayor limpieza y coherencia a toda la organización a través de la obtención de la máxima calidad del dato.



DATA MANAGEMENT

El gobierno del dato se encarga de la dirección y supervisión de los datos, una parte importante dentro de las áreas/fases del data management:

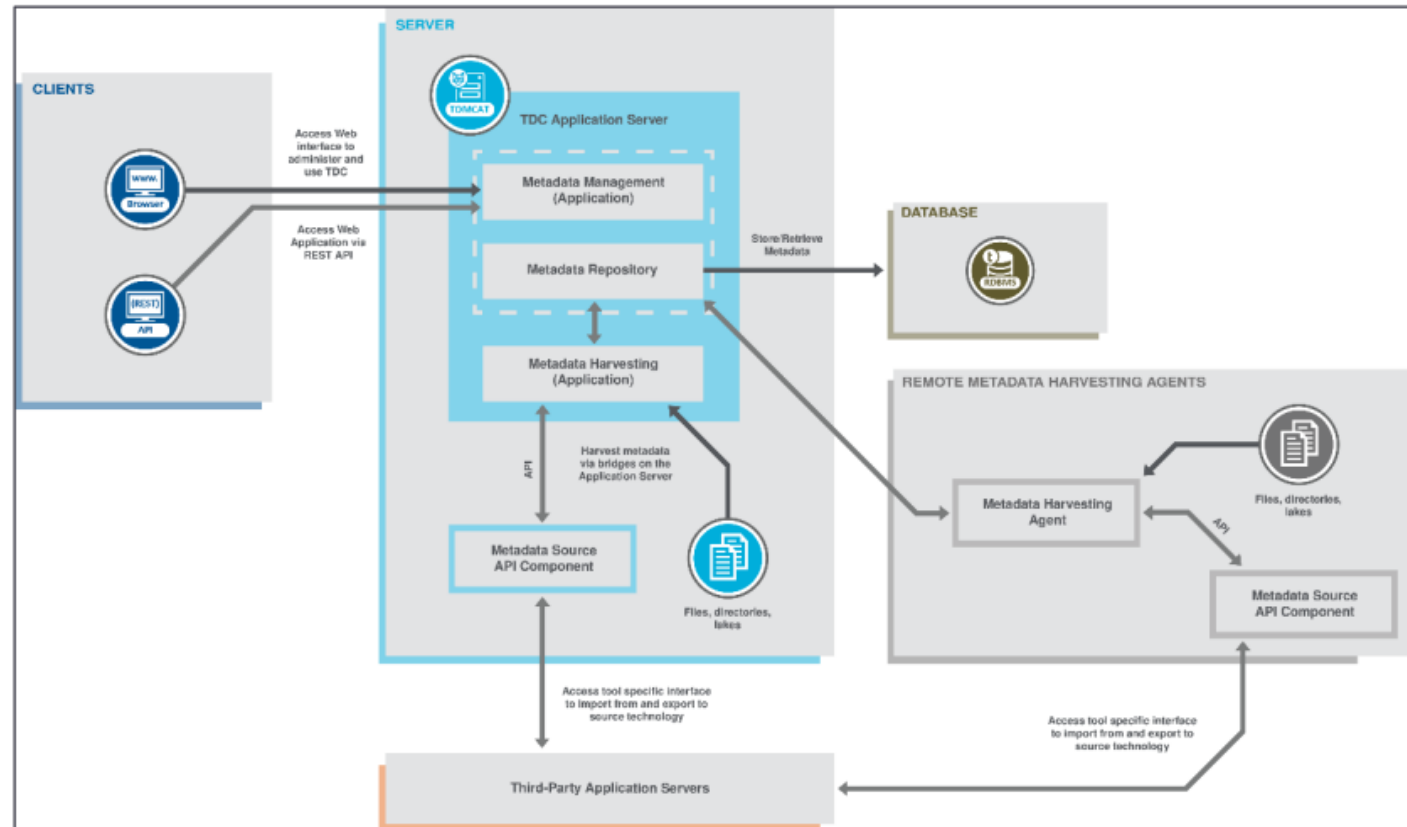
- Arquitectura de datos
- Modelización y diseño de datos
- Almacenamiento y operaciones de datos
- Seguridad de los datos
- Integración e interoperabilidad de datos
- Gestión de documentos y contenidos
- Datos de referencia y maestros
- Data Warehousing y BI
- Metadatos
- Calidad de los datos

TALEND DATA CATALOG



TALEND DATA CATALOG

Talend ofrece la herramienta **Talend Data Catalog** para realizar el proceso final y más abstracto de la gobernanza del dato a una organización. La gobernanza del dato no depende solamente de una plataforma o tecnología si no que es una serie de procesos orquestados durante el ciclo de vida del dato.



TALEND DATA CATALOG

Bloques de Talend Data Catalog:

- **Clients** (con un navegador WEB y una API REST).
- **Server** incluye la aplicación web de Talend Data Catalog:
 - *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.
- **Database** contiene la bdd utilizada para almacenar y recuperar metadatos, así como ejecutar cálculos sobre ellos (Oracle, PostgreSQL y MS SQL Server).
- Uno o más servidores de recolección de metadatos instalados en una máquina remota (**Remote Metadata Harvesting Agents**). Se utilizan para importar o exportar a la tecnología de la fuente de metadatos.

TALEND DATA CATALOG

Talend Data Catalog permite definir un catálogo central y gobernado de datos de confianza que puede ser compartido y elaborado de manera colaborativa fácilmente. Gracias a esta herramienta se puede descubrir, perfilar, organizar y documentar automáticamente los metadatos de toda una organización y hacer que sea fácil de buscar en ellos todo lo que se desee.

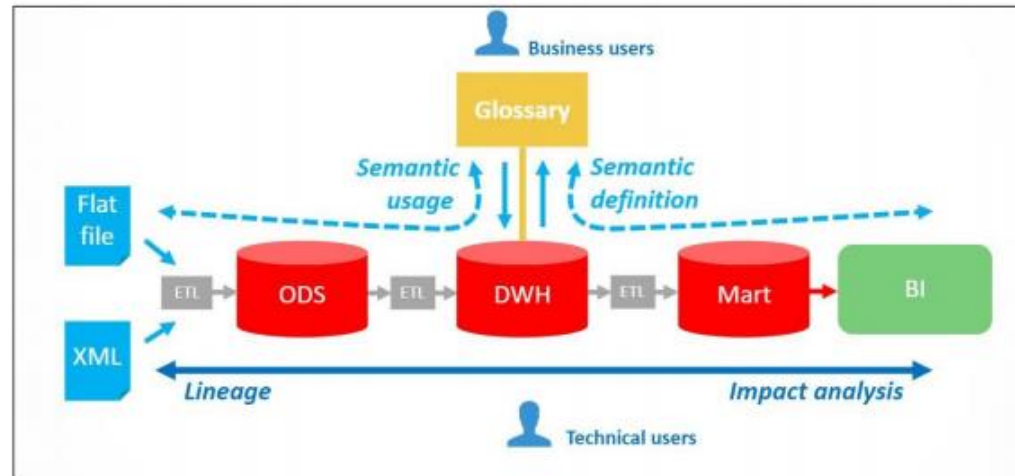
Sus principales características son:

- Análisis automático de datos (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.
- Validación y enmascaramiento de datos mediante diccionarios semánticos.
- Gestión de la seguridad de los archivos (roles y permisos predefinidos a usuarios y grupos).
- Social catalog es un conjunto de características que permiten el trabajo en equipo y facilitan la comunicación.

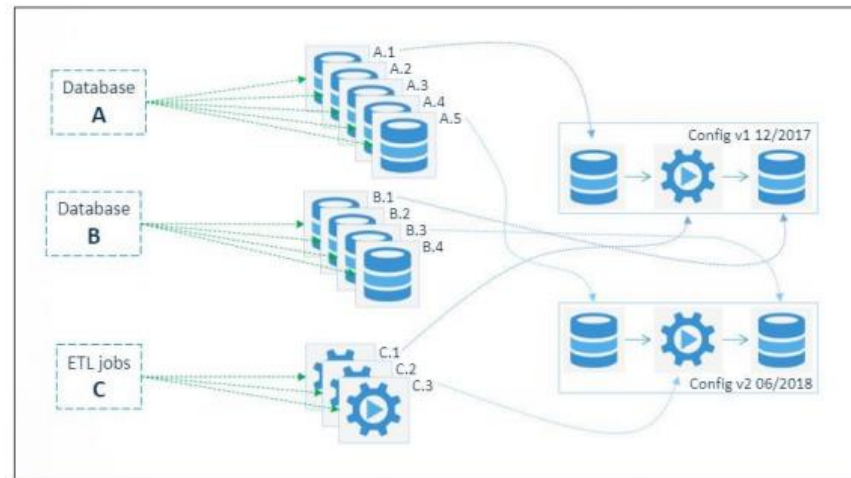


TALEND DATA CATALOG

- Gestión de los datos mediante glosarios que pueden ser trabajados con workflow.

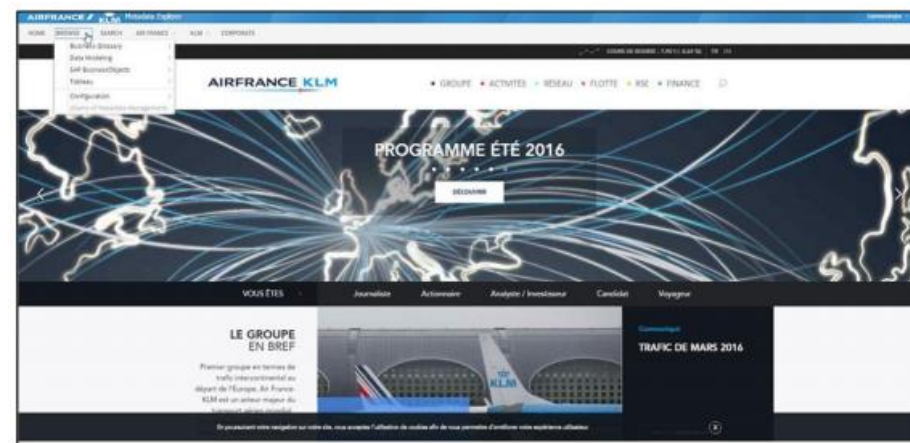
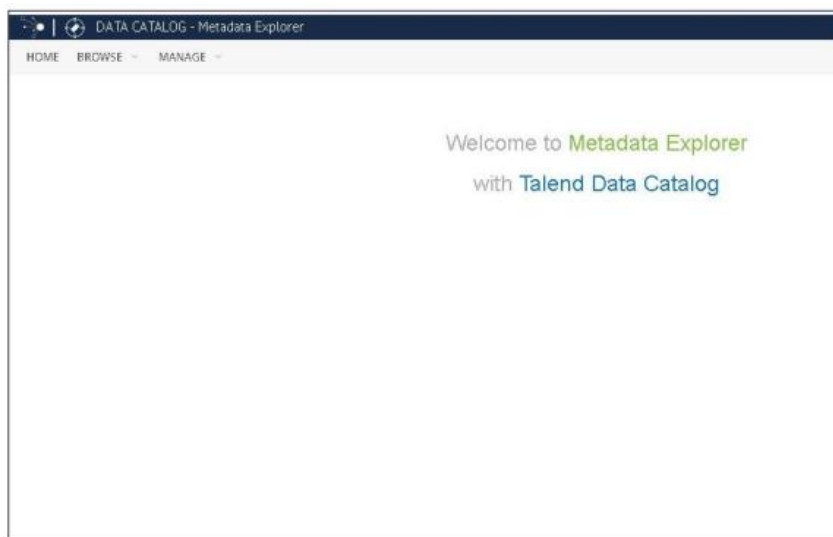


- Internamente trabaja con versiones de metadatos (por defecto ocultas).



TALEND DATA CATALOG

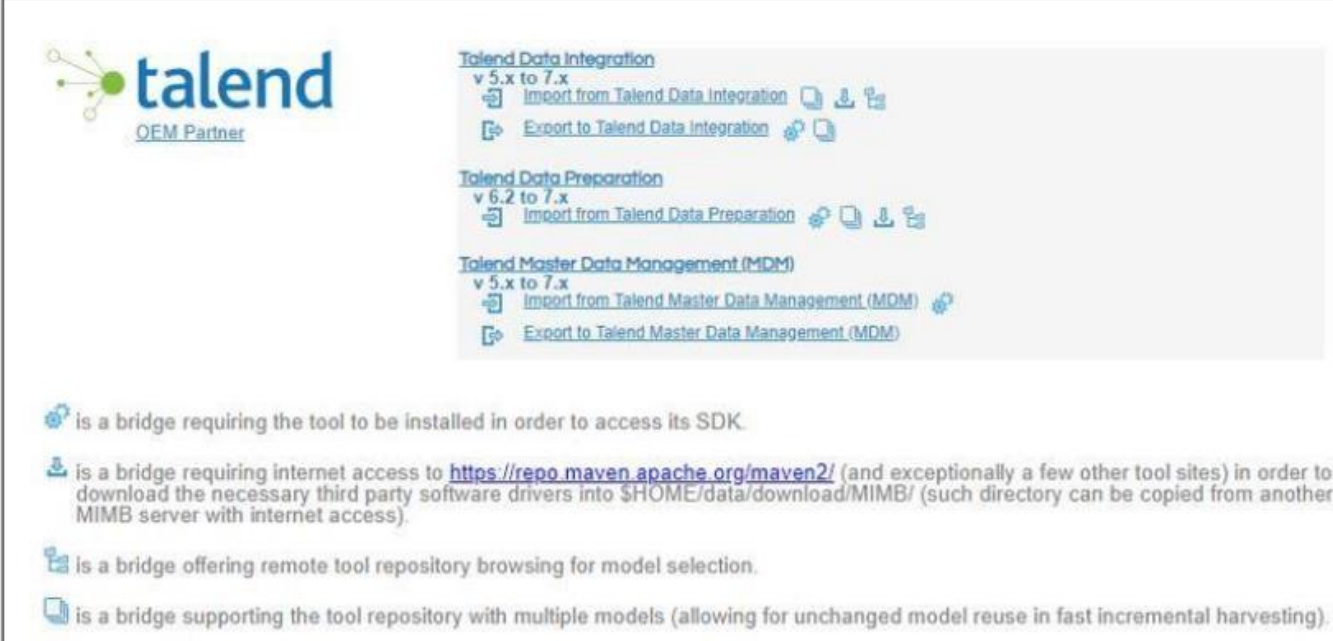
- La interfaz Metadata Explorer muestra las distintas configuraciones creadas en la interfaz Metadata Manager. Permite la modificación de los menús y el aspecto en función de los permisos que el usuario tenga.



- Metadata Manager es la interfaz de la herramienta donde se realiza la gestión completa tanto de los metadatos como de las vistas de los mismos. Permite la realización de copias de seguridad y versionado de los metadatos, así como la gestión de la seguridad de los mismos (roles/usuarios).

TALEND DATA CATALOG

Talend Data Catalog permite realizar conexiones con otras herramientas del entorno de Talend para poder realizar todas las fases de la gobernanza del dato de manera sencilla y organizada.



talend
OEM Partner

Talend Data Integration
v 5.x to 7.x
Import from Talend Data Integration
Export to Talend Data Integration

Talend Data Preparation
v 6.2 to 7.x
Import from Talend Data Preparation

Talend Master Data Management (MDM)
v 5.x to 7.x
Import from Talend Master Data Management (MDM)
Export to Talend Master Data Management (MDM)

is a bridge requiring the tool to be installed in order to access its SDK.

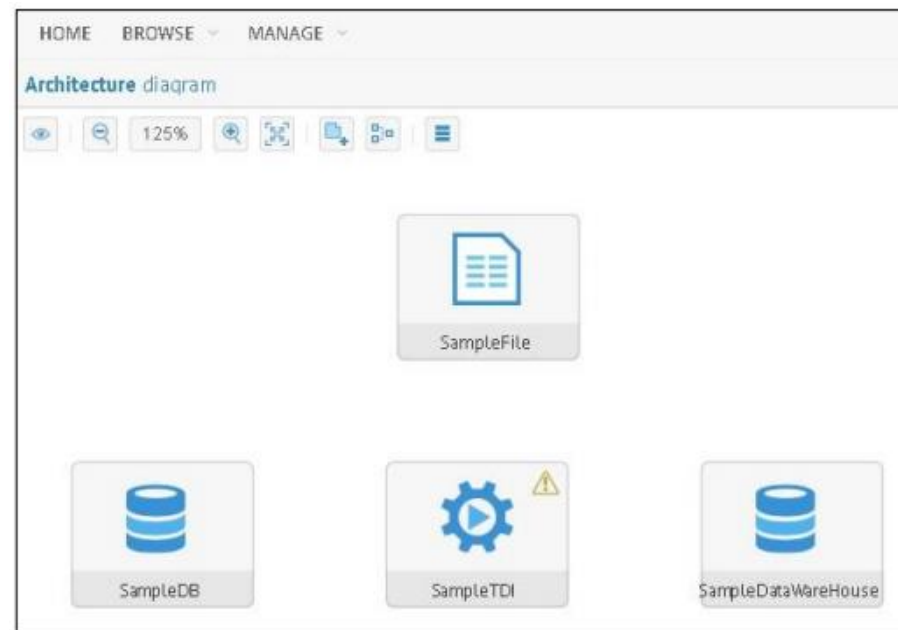
is a bridge requiring internet access to <https://repo.maven.apache.org/maven2/> (and exceptionally a few other tool sites) in order to download the necessary third party software drivers into \$HOME/data/download/MIMB/ (such directory can be copied from another MIMB server with internet access).

is a bridge offering remote tool repository browsing for model selection.

is a bridge supporting the tool repository with multiple models (allowing for unchanged model reuse in fast incremental harvesting).

TALEND DATA CATALOG

STITCHING METADATA es una funcionalidad de la herramienta que detecta ambigüedades que no pueden ser resueltas automáticamente en los metadatos. Se realiza un proceso donde se crea una relación entre modelos para poder analizar el impacto y el linaje de los datos.



TALEND DATA PREPARATION & STEWARDSHIP



TALEND DATA PREPARATION

Talend Data Preparation (TDP) es una aplicación en la nube que permite simplificar y acelerar el proceso de preparación de datos para realizar posteriormente el resto de fases de un proceso ETL.

Presentación intuitiva de las diferentes funcionalidades a través de su interfaz.

The screenshot displays the Talend Data Preparation (TDP) interface. At the top, there's a navigation bar with 'Data Preparation' and 'Stratebi Business Solutions'. Below this, the main header shows 'sandyrelated Preparation' and 'Dataset: sandyrelated'. The left sidebar contains a '1 Change date format on column Closed Date' section with a 'SUBMIT' button. The main area shows a table with 8 rows and 9 columns: Unique Key (integer), Created Date (date), Closed Date (date), Agency (text), Agency Name (text), Complaint Type (text), and Descriptor (text). The table is filtered to show 10000/10000 rows. On the right, a '2 columns selected' panel shows a list of columns and a search filter.

	Unique Key integer	Created Date date	Closed Date date	Agency text	Agency Name text	Complaint Type text	Descriptor text
1	24317811	10/28/2012 12:00:...	2012-11-08	HPD	Department of Hou...	HEATING	HEAT
2	24317782	10/28/2012 12:00:...	2012-11-10	HPD	Department of Hou...	NONCONST	VERMIN
3	24317758	10/28/2012 12:00:...		HPD	Department of Hou...	HEATING	HEAT
4	24317645	10/28/2012 12:00:...	2012-11-01	HPD	Department of Hou...	HEATING	HEAT
5	24317640	10/28/2012 12:00:...	2012-11-01	HPD	Department of Hou...	HEATING	HEAT
6	24317636	10/28/2012 12:00:...	2012-10-29	HPD	Department of Hou...	HEATING	HEAT
7	24317634	10/28/2012 12:00:...	2012-11-02	HPD	Department of Hou...	HEATING	HEAT
8	24317630	10/28/2012 12:00:...	2012-10-29	HPD	Department of Hou...	HEATING	HEAT

Esta herramienta trabaja sobre el conjunto de datos sin realizar cambios sobre los datos originales. Es decir, partiendo del conjunto de datos importado (dataset), se realizan cambios que crean un nuevo conjunto de datos (preparación/receta) que permite ser exportado sin modificar el original (xls, csv y otros formatos). Las recetas son reutilizables para aplicarlas en otros conjuntos de datos de características similares

TALEND DATA PREPARATION

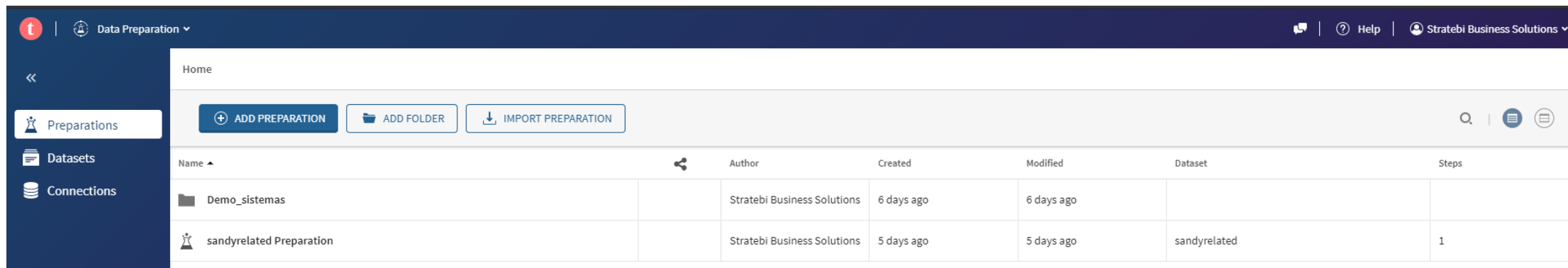
Las características de esta herramienta de Talend son:

- Integración y catalogación de los datos.
- Descubrimiento y creación de perfiles de datos (para el futuro análisis).
- Limpiar, estandarizar y dar forma al conjunto de datos en función de los requisitos buscados/establecidos.
- Enriquecimiento y conexión de conjuntos de datos.
- Puesta en funcionamiento de la preparación de datos para su posterior uso.

TALEND DATA PREPARATION

Algunas funcionalidades de la transformación y limpieza del dato son:

- El reemplazo de valores.
- La concatenación o división de columnas.
- La elección del formato del campo.
- El análisis de valores nulos o no válidos.
- El uso y aplicación de operaciones matemáticas avanzadas sobre los campos.
- Cruzado de datos entre tablas de distintos datasets.



The screenshot displays the Talend Data Preparation web interface. On the left is a dark blue sidebar with navigation links: Preparations (selected), Datasets, and Connections. The main area has a top bar with 'Data Preparation' and user information. Below this is a 'Home' section with three buttons: 'ADD PREPARATION', 'ADD FOLDER', and 'IMPORT PREPARATION'. A table lists existing preparations with columns for Name, Author, Created, Modified, Dataset, and Steps.

Name	Author	Created	Modified	Dataset	Steps
Demo_sistemas	Stratebi Business Solutions	6 days ago	6 days ago		
sandyrelated Preparation	Stratebi Business Solutions	5 days ago	5 days ago	sandyrelated	1

TALEND DATA STEWARDSHIP

Talend Data Stewardship (TDS) permite validar y resolver conflictos en los datos de forma colaborativa, así como abordar posibles problemas de integridad de los datos. Organiza las interacciones sobre los datos (siempre que se necesite la intervención humana) para colaborar en la conservación, el arbitraje o la validación de los datos.

The screenshot displays the Talend Data Stewardship (TDS) web interface. The top navigation bar includes the Talend logo, a user profile icon, and the text 'Data Stewardship'. On the right, there are links for 'Help' and 'Stratebi BUSINESS SOLUTIONS'. A left sidebar contains navigation options: 'Tasks' (selected), 'Campaigns', and 'Data models'. The main content area shows a list of tasks with the following details:

Task Name	Description	Type	Progress	Start date	My tasks	Unassigned	Data model
Demo - CRM data deduplication - sistemas@nfr.stratebi.com	Identify unique customers based on Salesforce, Marketo and Netsuite data.	Merging	50 %	2020-03-23	20	20	Demo - Customers
Demo - Find beta tester candidates - sistemas@nfr.stratebi.com	Identify the best candidates for beta testing among all vendors.	Arbitration	50 %	2020-03-23	20	20	Demo - Customers

At the top right of the task list, there are buttons for 'ARBITRATION', 'RESOLUTION', 'MERGING', and 'GROUPING'. The interface also includes search and sorting controls at the top of the task list.

TALEND DATA STEWARDSHIP

La herramienta trabaja mediante *Campañas* que permiten organizar las distintas *tareas* a cumplir. Un administrador decide a que usuario asignar cada tarea.

The screenshot displays the Talend Data Stewardship interface for a campaign titled "Demo - CRM data deduplication - sistemas@nfr.stratebi.com". The interface includes a sidebar for assigning tasks, a main table of data records, and a right-hand panel with a search bar and a chart.

Table Data:

ID	First_Name	Last_Name	Gender	Age	Occupation	Company	Address	City	State
1	Ryan	Butt	F	15	K-12 Student	Greentaxon	6649 N Blue G...	New Orleans	LA
2	Ernest	Darakjy	M	62	Self-Employed	Ap-lux	4 B Blue Ridg...	Brighton	MI
3	Steven	Venere	M	28	Scientist	Move-it	8 W Cerritos ...	Bridgeport	NJ
4	Theophile	Paprocki	M	45	Executive/Man...	Superox	639 Main St	Anchorage	AK
5	Evan	Foller	M	31	Writer	Lotfind	34 Center St	Hamilton	OH
6	Cyril	Morasca	M	53	Unemployed	Tampity	3 Mcauley Dr	Ashland	OH
7	Onar	Tollner	M	41	Academic/Educ...	Groove-is	7 Eads St	Chicago	IL
8	Oscar	Dilliard	M	32	Programmer	keyzin	7 W Jackson Blvd	San Jose	CA
9	Lena	Wieser	F	29	Technical/Eng...	Singlezone	5 Boston Ave #88	Sioux Falls	SD
10	Sarah	Harrier	F	42	Academic/Educ...	Strongcode	228 Runamuck ...	Baltimore	MD
11	Ryan	Butt	F	15	K-12 Student	Greentaxon	6649 N Blue G...	New Orleans	LA
12	Ernest	Darakjy	M	62	Self-Employed	Ap-lux	4 B Blue Ridg...	Brighton	MI
13	Steven	Venere	M	28	Scientist	Move-it	8 W Cerritos ...	Bridgeport	NJ
14	Theophile	Paprocki	M	45	Executive/Man...	Superox	639 Main St	Anchorage	AK
15	Evan	Foller	M	31	Writer	Lotfind	34 Center St	Hamilton	OH
16	Cyril	Morasca	M	53	Unemployed	Tampity	3 Mcauley Dr	Ashland	OH
17	Onar	Tollner	M	41	Academic/Educ...	Groove-is	7 Eads St	Chicago	IL

TALEND DATA STEWARDSHIP

Cada campaña permite decidir el tipo de campaña a resolver, los involucrados, la estructura de datos utilizada y las tareas necesarias para completar cada campaña. Es una herramienta colaborativa ya que los usuarios pueden explorar los datos relacionados con sus tareas, resolver las tareas individualmente, delegar tareas a otros compañeros y monitorear y auditar las tareas de otros compañeros. A su vez, existe la posibilidad de activar flujos de trabajo de validación para tareas que deben verificarse dos veces.

The screenshot displays the 'Add campaign' interface in the Talend Data Stewardship application. The interface features a dark blue sidebar on the left with a navigation menu containing four items: 'General' (highlighted with a yellow circle and the number 1), 'Roles' (with a blue circle and the number 2), 'Data model' (with a blue circle and the number 3), and 'Workflow' (with a blue circle and the number 4). The main content area is divided into two sections. The top section, titled '1/ General', contains a 'NAME' text field, a 'DESCRIPTION (optional)' text area, a 'TYPE' dropdown menu with the placeholder text 'Select a campaign type', and a toggle switch for 'ENABLE TASK RESOLUTION DELAY'. Below these fields is a 'Campaign owners' section with a '+ Add a campaign owner' button and a tag labeled 'Stratebi Business Solutions'. The bottom section, titled '2/ Roles', includes a 'Stewards' checkbox that is checked, a '+ Add a steward' button, and an 'ADD A ROLE' button at the bottom. The top of the interface has a dark blue header bar with a 'Data Stewardship' dropdown menu, a 'Help' icon, and a 'Stratebi BUSINESS SOLUTIONS' dropdown menu.

TALEND DATA STEWARDSHIP

Esta herramienta permite a los usuarios (no necesariamente expertos o profesionales de datos), administrar y resolver rápidamente cualquier problema de integridad de los datos para lograr datos "confiables" en toda la empresa de una sencilla manera.

En Talend Data Stewardship:

- Se definen modelos de datos comunes, semánticas y reglas necesarias para limpiar y validar datos.
- Se definen roles de usuario, flujos de trabajo y prioridades.
- Las tareas se delegan a las personas que mejor conocen los datos.
- Se resuelven tareas y campañas para lograr la máxima calidad del dato y con ello, el éxito.

Preguntas

