

# Data Management:

## Gestión holística de los datos empresariales



**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.

# Data Management:

Gestión holística de los datos empresariales

## ARQUITECTURA DE DATOS



**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.

# Arquitectura de Datos

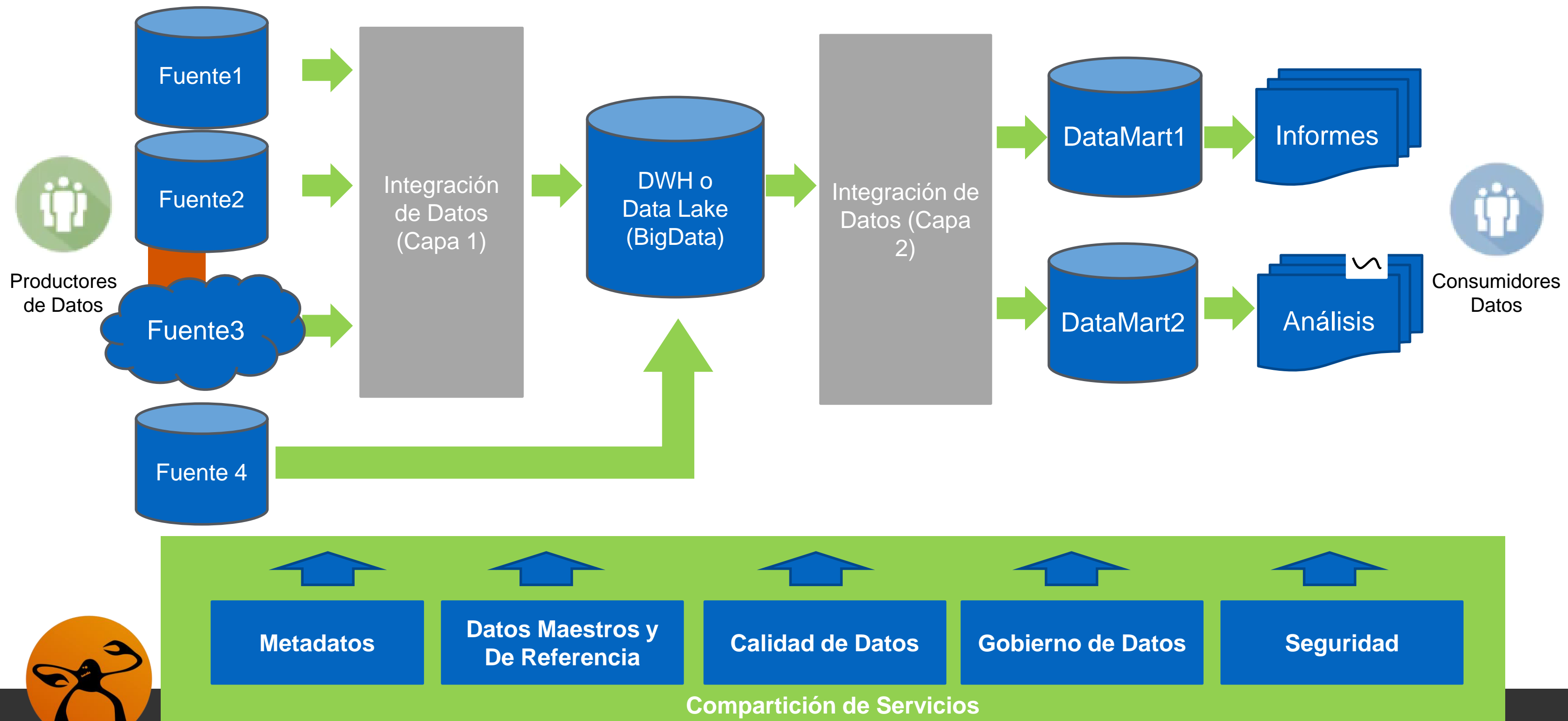
El **Arquitectura de Datos** se refiere a los modelos, políticas, reglas o estándares que rigen qué datos se recopilan, y cómo se almacenan, se organizan y se utilizan en los sistemas de una organización.



En cuanto a las funciones de gestión de datos, la arquitectura de datos cubre cómo cada función encaja en el marco general de gestión de datos.



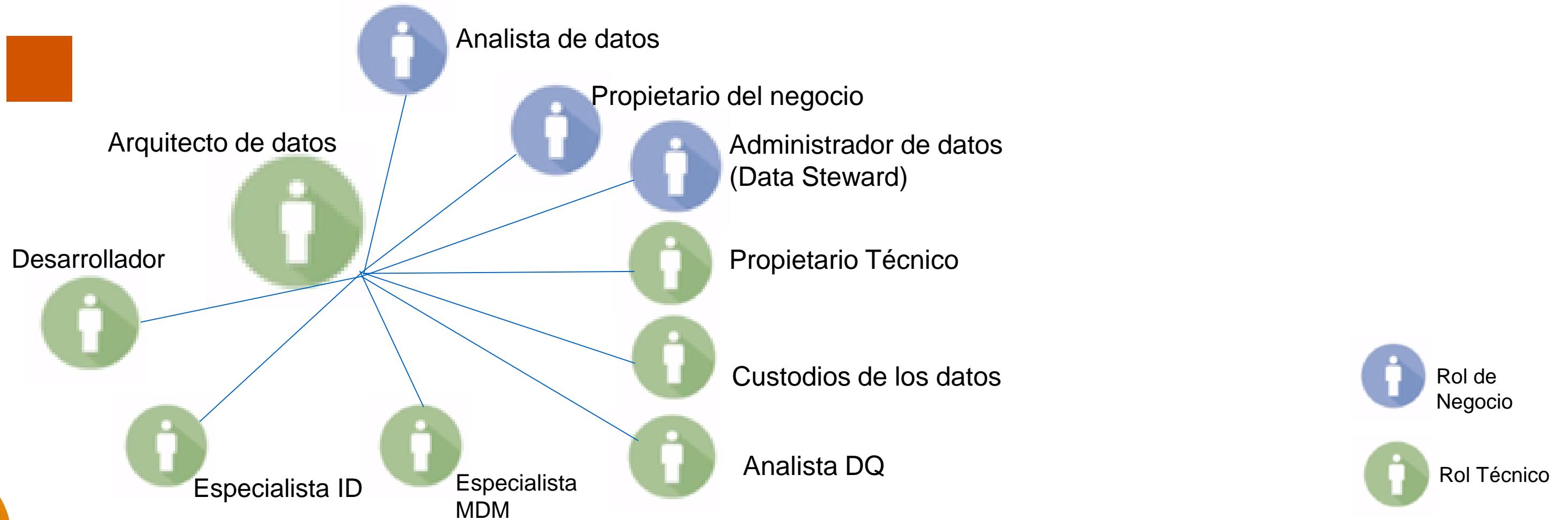
# Ejemplo: Arquitectura Analítica de Datos



# Roles de Arquitectura de datos



El **Arquitecto de datos** es responsable principal de diseñar la arquitectura de datos. Aunque este rol es específico de la arquitectura de datos, el arquitecto de datos trabajará estrechamente con todos los demás roles de administración de datos. Las responsabilidades principales de este rol incluyen: el diseño y la optimización de la arquitectura de datos a través de las capas de arquitectura de datos. También los arquitectos de datos proponen las tecnologías necesarias para dar soporte a la arquitectura empresarial y las funciones de la gestión de datos.



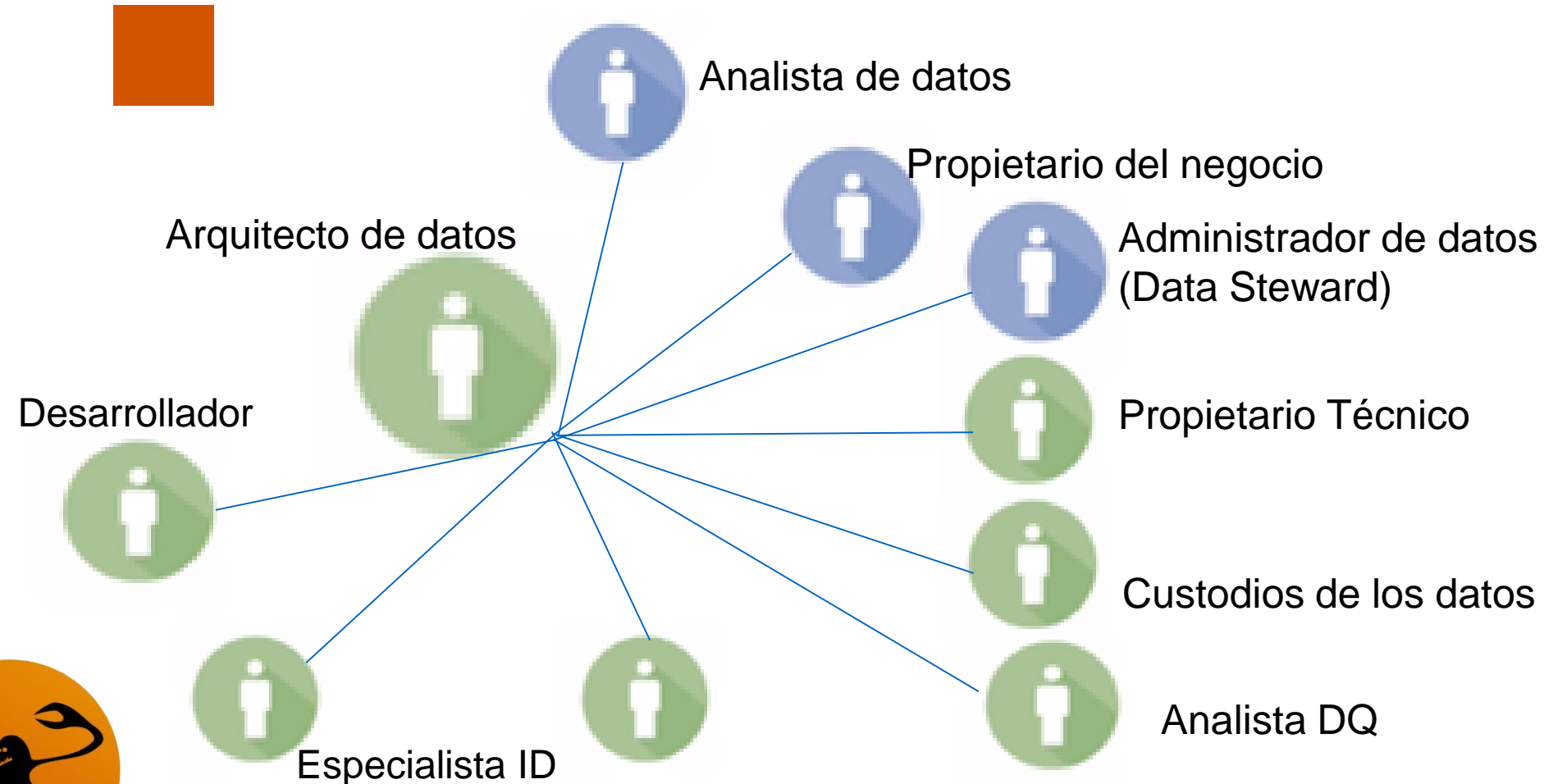
# Rol de Arquitectura de datos



El **Arquitecto de datos** es responsable principalmente de diseñar la arquitectura de datos. Aunque este rol es específico de la arquitectura de datos, el arquitecto de datos trabajará estrechamente con todos los demás roles de administración de datos. Las responsabilidades principales de este rol incluyen: el diseño y la optimización de la arquitectura de datos a través de las capas de arquitectura de datos. También los arquitectos de datos proponen las tecnologías necesarias para dar soporte a la arquitectura empresarial y las funciones de la gestión de datos.




## PERSONAS



*La definición del Rol en la Arquitectura de Datos y sus responsabilidades cubren la perspectiva de personas dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management*




# Resumen – Puntos Clave



La **Arquitectura de Datos** hace referencia a los **modelos, políticas, reglas o estándares** que rigen qué datos se recopilan, y cómo se almacenan, organizan y ponen en uso en un sistema de base de datos y / o en una organización.



El **Rol** principal en la Arquitectura de Datos es el **Arquitecto de Datos**



La Arquitectura de Datos está compuesta de un conjunto de capas, llamadas **funciones de la gestión de datos**. La arquitectura de datos cubre cómo función de la gestión de datos, encaja en el marco general de gestión de datos.



# Data Management:

Gestión holística de los datos empresariales

## INTEGRACIÓN DE DATOS



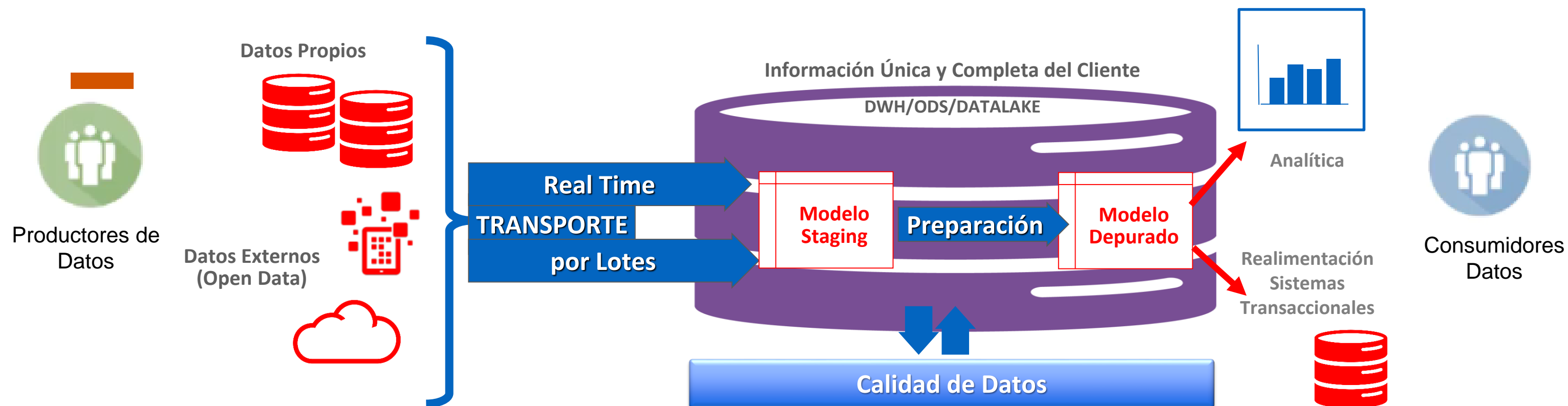
**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.



# Gestión de Integración de Datos

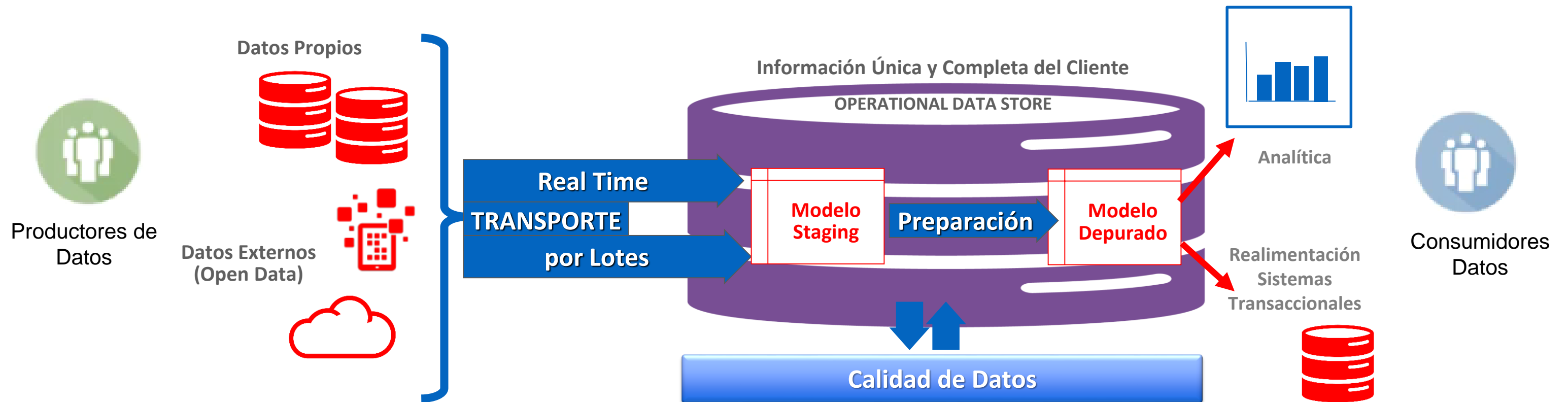
La **Integración de Datos** hace referencia al proceso y tecnologías que permiten el movimiento de datos desde los sistemas de datos de origen a los sistemas de datos destino. En su camino, los datos se transforman en información para adaptarse a los requisitos del negocio.

DE CUALQUIER FUENTE CUANDO SE NECESITE CON LAS CARACTERISTICAS QUE SE NECESITEN



# Escenarios de Integración de Datos - ETL

ETL (Extracción Transformación Carga) hace referencia al enfoque de integración de datos donde se **extraen** los datos de los Sistemas de Origen, luego pasan por el proceso de **transformación** y termina **cargando** en el Sistema de Destino. El escenario ETL es típico utilizado en sistemas DataWareHouse.



**E:** Extracción de datos desde Fuentes Origen

**T:** Transformación de datos

**L:** Carga de datos transformados a Fuentes Destino

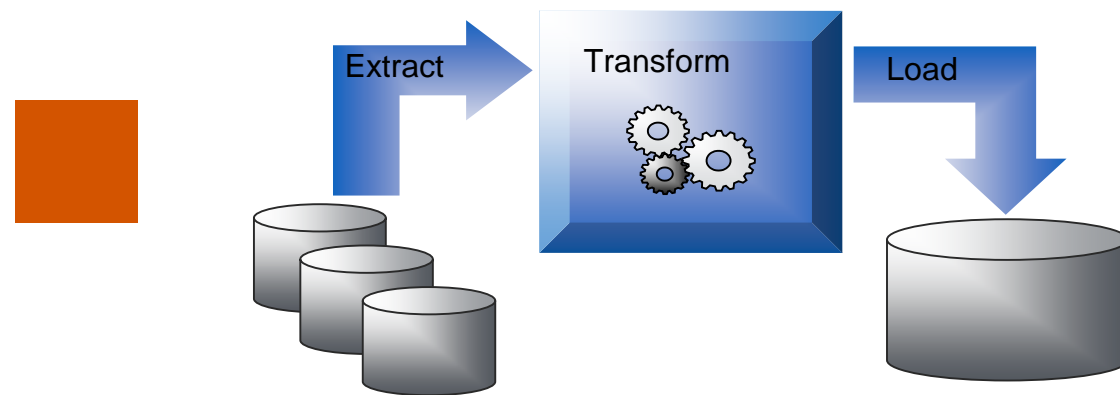
*Típicos Ejemplos de Transformación :  
Cálculos y Agregaciones*



# Escenarios de Integración de Datos . EL-T

**EL-T (Extracción Carga Transformación)** EL-T hace referencia al enfoque de integración de datos donde los datos se extraen de los sistemas de origen y luego se cargan en el sistema de destino sin ninguna transformación. La transformación de los datos se realiza más tarde en el sistema de destino. El escenario ELT es típico de los sistemas Big Data/Hadoop.

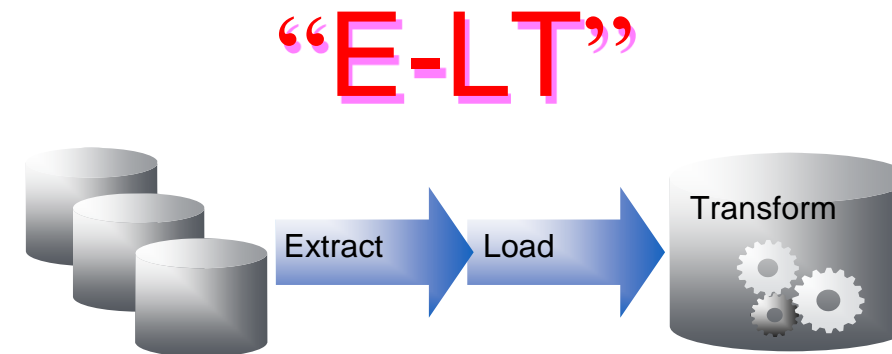
Arquitectura ETL Tradicional



Modelo ETL Tradicional

- Menor Rendimiento
- Mayor Coste – HW de transformación

ODI - Arquitectura E-LT



Modelo E-LT

- Mayor Rendimiento – 1 única carga
- Sin Costes adicionales en HW



# Integración de Datos. Batch vs. Tiempo Real

## Batch

En el **Procesamiento Batch** (por lotes), se recopilan un gran grupo de transacciones y los datos se procesan durante una única ejecución.

### Algunas Consideraciones:

- Debido al gran volumen de datos, el proceso debe ejecutarse cuando los recursos están menos ocupados, generalmente durante la noche.
- El procesamiento batch retrasa el acceso a los datos, requiere una estrecha supervisión y existe la posibilidad de tiempo de inactividad.
- Debido al retraso al acceso a los datos, se pierde el conocimiento hasta que termine el procesamiento.
- Los problemas que ocurren durante el procesamiento batch pueden demorar todo el proceso, por lo que necesita personal de soporte para monitorizarlo desde cerca mientras se ejecuta.

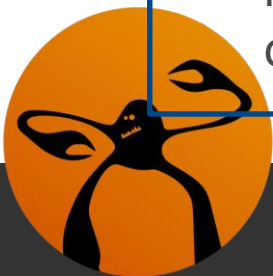
## Real Time

En el **Procesamiento en Tiempo Real**, se procesan pequeños grupos de transacciones a demanda.

### Algunas Consideraciones:

- La ventaja del procesamiento en tiempo real es que brinda acceso instantáneo a las ejecuciones de datos con menos recursos y mejora el tiempo de actividad.
- Con la integración de datos en tiempo real, se tiene conocimiento del negocio a medida que ocurren las transacciones.
- Si ocurre un error, puede ser manejado inmediatamente.
- Es más complejo el diseño del procesamiento en tiempo real.

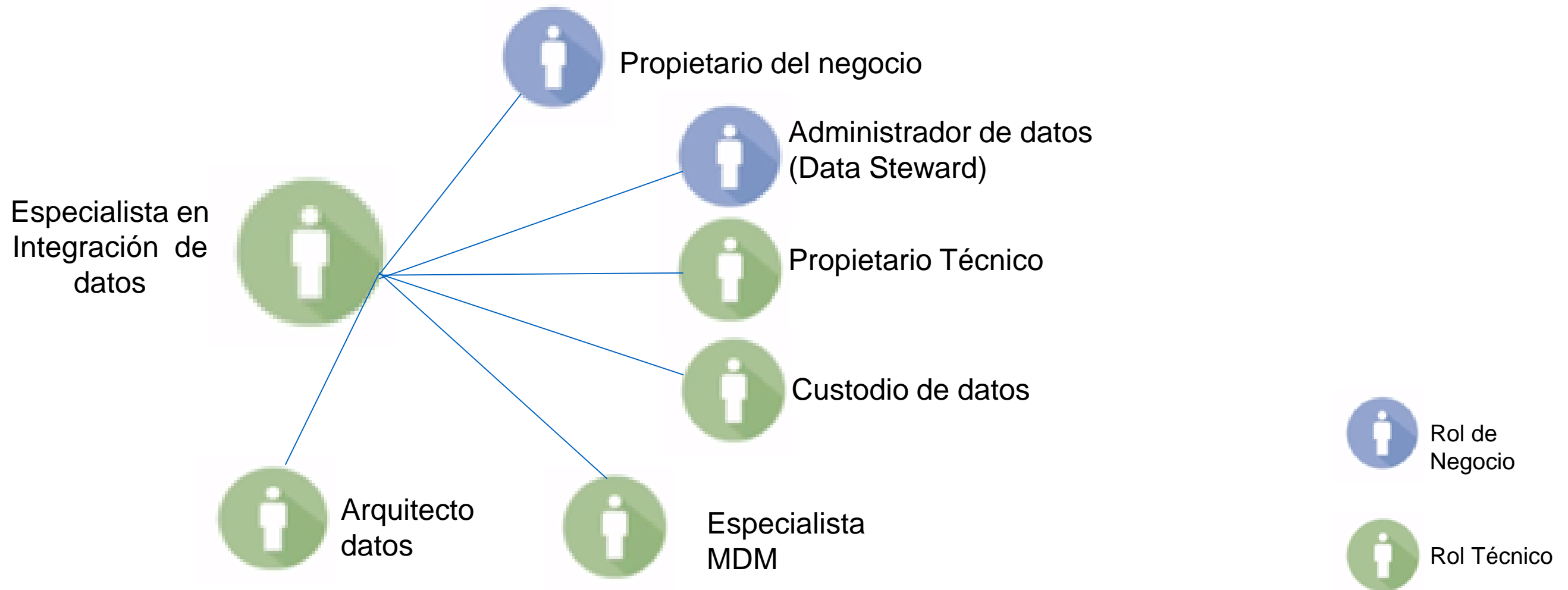
Aunque los sistemas con procesamiento en tiempo real requieren más esfuerzo para diseñar e implementar, el beneficio para el negocio puede ser enorme.



# Rol de Integración de datos



El **Especialista en Integración de Datos** es responsable principal de realizar las actividades asociadas con la Integración de Datos. Los especialistas en integración de datos trabajarán estrechamente con los propietarios de negocios, administradores de datos, propietarios técnicos, custodios de datos, especialistas en MDM y arquitectos de datos. Sus responsabilidades principales se refieren al diseño e implementación de aplicaciones de integración de datos, incluida la definición de especificaciones de mapeo, diseño e implementación de Jobs de integración de datos, etc



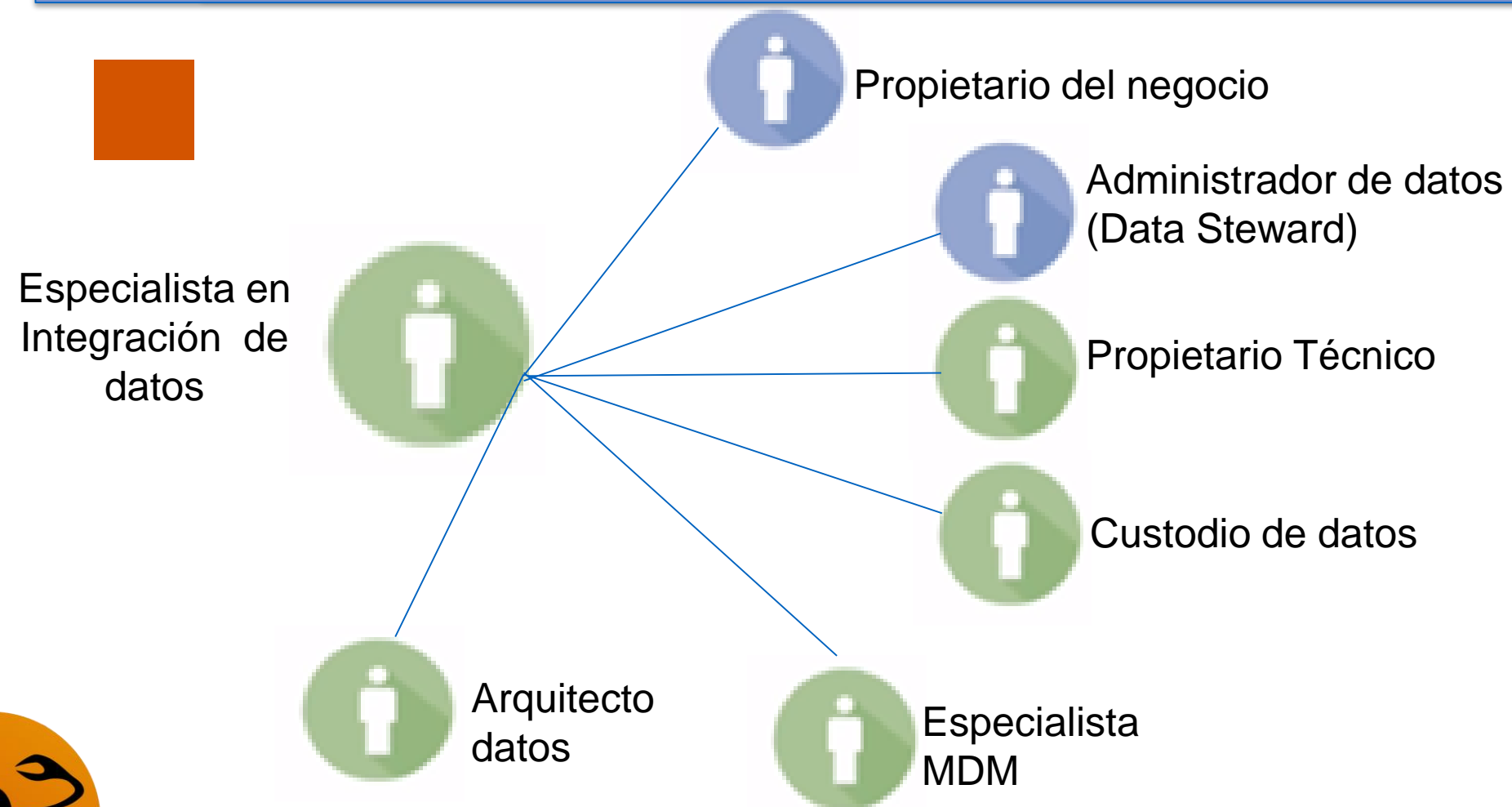
# Rol de Integración de datos



El **Especialista en Integración de Datos** es responsable principal de realizar las actividades asociadas con la Integración de Datos. Los especialistas en integración de datos trabajarán estrechamente con los propietarios de negocios, administradores de datos, propietarios de datos, custodios de datos, especialistas en MDM y arquitectos de datos. Sus responsabilidades principales se refieren al diseño e implementación de aplicaciones de integración de datos, incluida la definición de especificaciones, el diseño e implementación de Jobs de integración de datos, etc



## PERSONAS



*La definición del Rol en la Arquitectura de Datos y sus responsabilidades cubren la perspectiva de personas dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management*



# Herramientas para Integración de Datos

## Las características principales de las herramientas de integración de datos.

- Capacidad para realizar procesamiento batch y en tiempo real.
- Capacidad realizar detección de cambio de datos (Identificar registros modificados).
- Capacidad de para realizar potentes transformaciones de datos estructurados y no estructurados.
- Capacidad para operaciones de manejo de errores integrales.

## HERRAMIENTAS DE INTEGRACIÓN DE DATOS

PROCESAMIENTO  
TIEMPO REAL

PROCESAMIENTO  
BATCH

DETECCIÓN DE  
CAMBIOS DE  
DATOS (CDC)

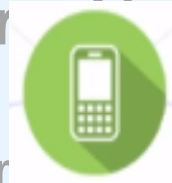
TRANSFORMACION  
ES





# Herramientas para Integración de Datos

Las características principales de las herramientas de integración de datos.



- Capacidad para realizar procesamiento batch y en tiempo real.
- Capacidad para realizar detección de cambio de datos (Identificar registros modificados).
- Capacidad de para realizar potentes transformaciones de datos estructurados y no estructurados.
- Capacidad para operaciones de manejo de errores integrales.

## TECNOLOGÍA

### HERRAMIENTAS DE INTEGRACIÓN DE DATOS

PROCESAMIENTO  
TIEMPO REAL

PROCESAMIENTO  
BATCH

DETECCIÓN DE  
CAMBIOS DE  
DATOS (CDC)


TRANSFORMACIONES

*Las Herramientas Tecnológicas que dan soporte a la integración de datos hacen referencia a la **perspectiva de tecnología** dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management*







# Resumen – Puntos Clave




La **Integración de Datos** hace referencia al movimiento de datos desde los Productores de datos (Fuentes Origen) a los consumidores de datos (Fuentes Destino).




El **RoI** principal en la Integración de Datos es el **Especialista en Integración de Datos**



Los escenarios de integración de datos más comunes son, **ETL (Extracción, Transformación y Carga)**, y el **EL-T (Extracción, Carga y Transformación)**.



Las características principales de las **Herramientas de integración de datos** están relacionadas con la capacidad de realizar procesamiento batch y en tiempo real, detección de cambios en datos y el manejo de errores.



La integración de datos puede basarse en **procesos batch (por lotes) o en tiempo real**. El procesamiento en tiempo real es común en los entornos de big data.



# Data Management:

## Gestión holística de los datos empresariales

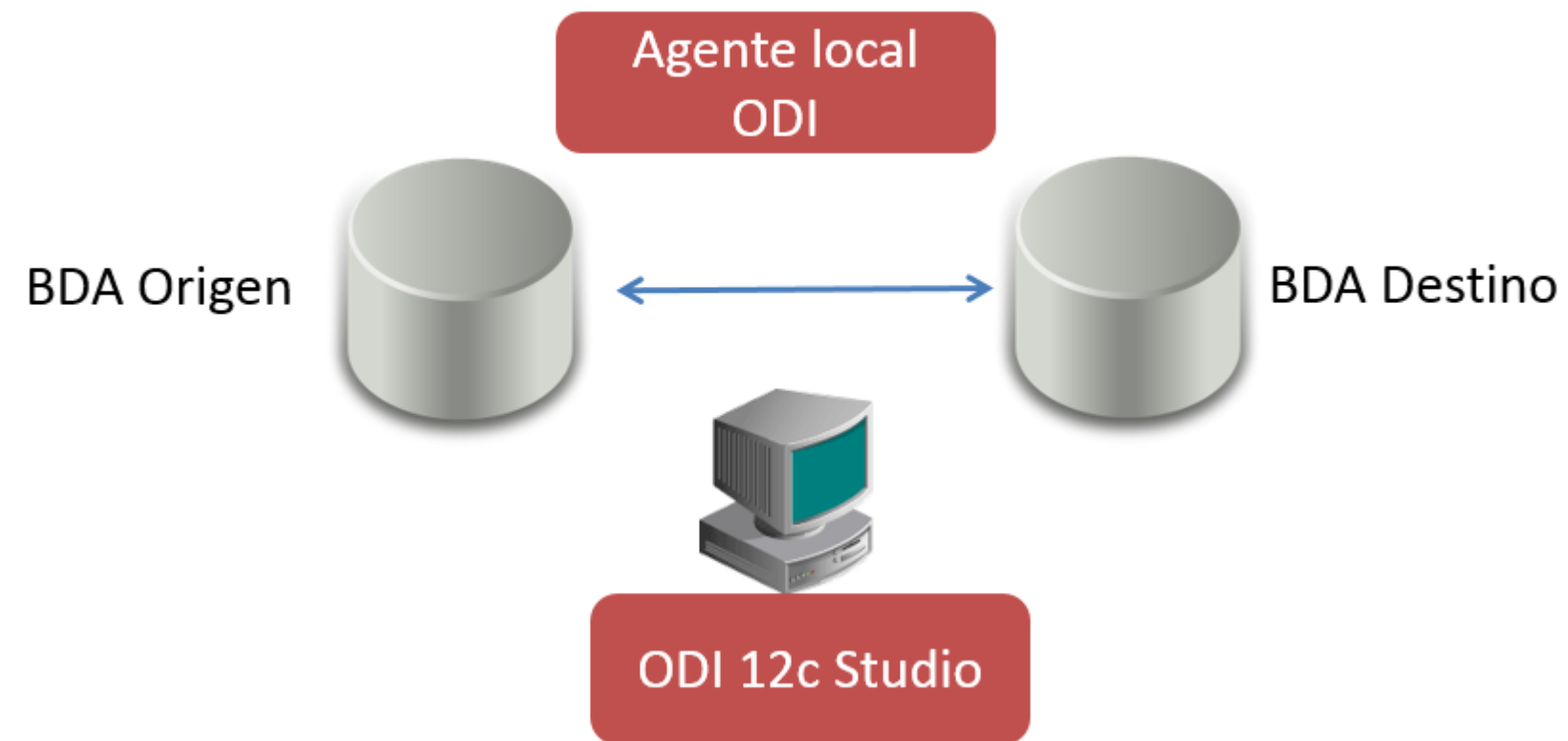
# Ejercicios Integración



**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.

# Ejercicios Integración

## Planteamiento Masterclass



# Ejercicios Integración



## PARTE I . ORACLE DATA INTEGRATOR

- Creación de Repositorio Maestro y de Trabajo
- Conexión al Repositorio de Trabajo ODI12c.
- Ejercicios de Integración de datos

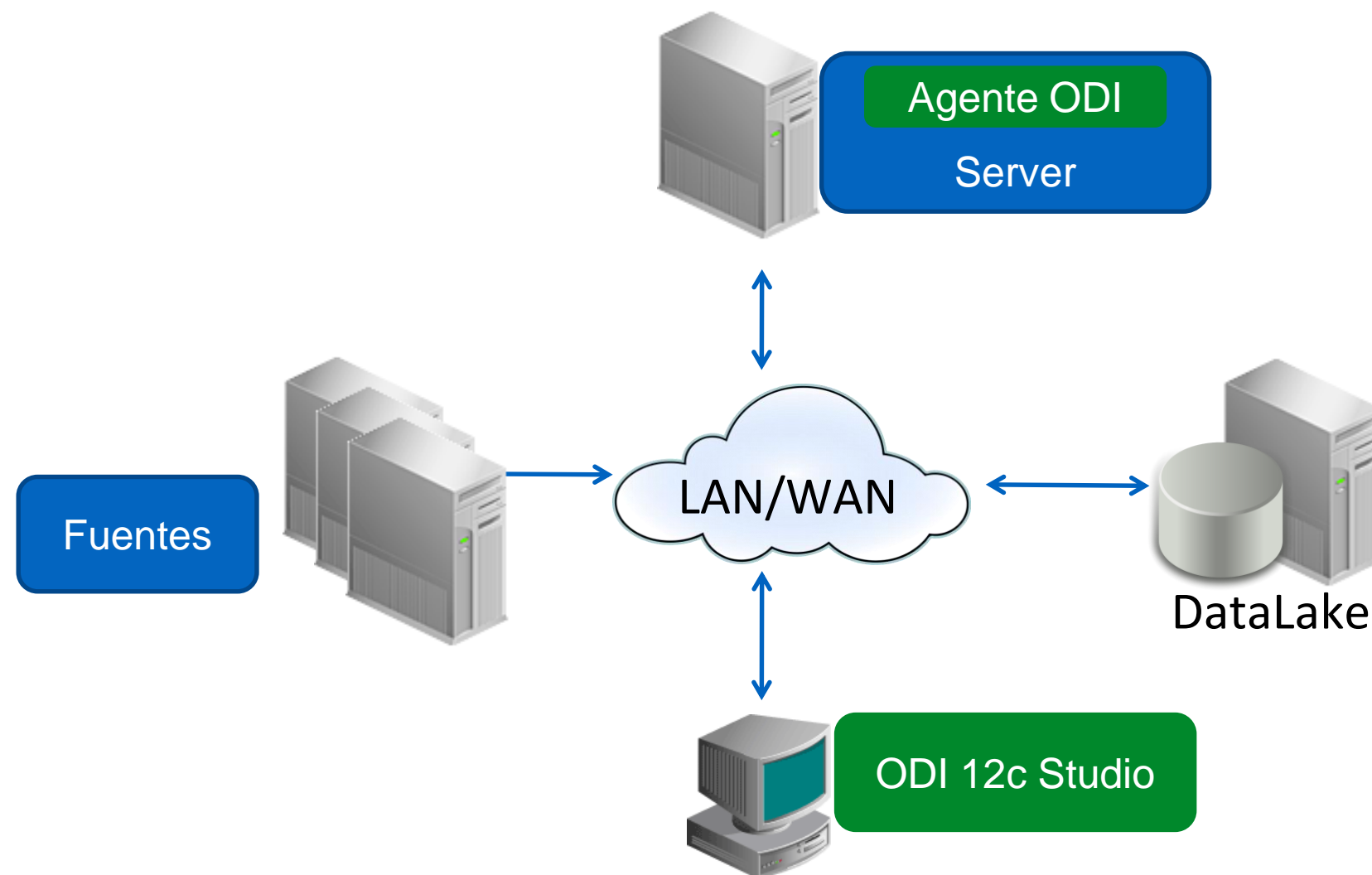
## PARTE II . TALEND DATA INTEGRATION

- Creación de Conexiones
- Ejercicios de Integración de datos



# Ejercicios Integración

## Planteamiento Real



# Ejercicios Integración



## Contextos

En ODI se utilizan los contextos para definir el entorno donde se quieren ejecutar los proyectos

- Localmente
- Desarrollo
- Producción
- Test/PreProducción

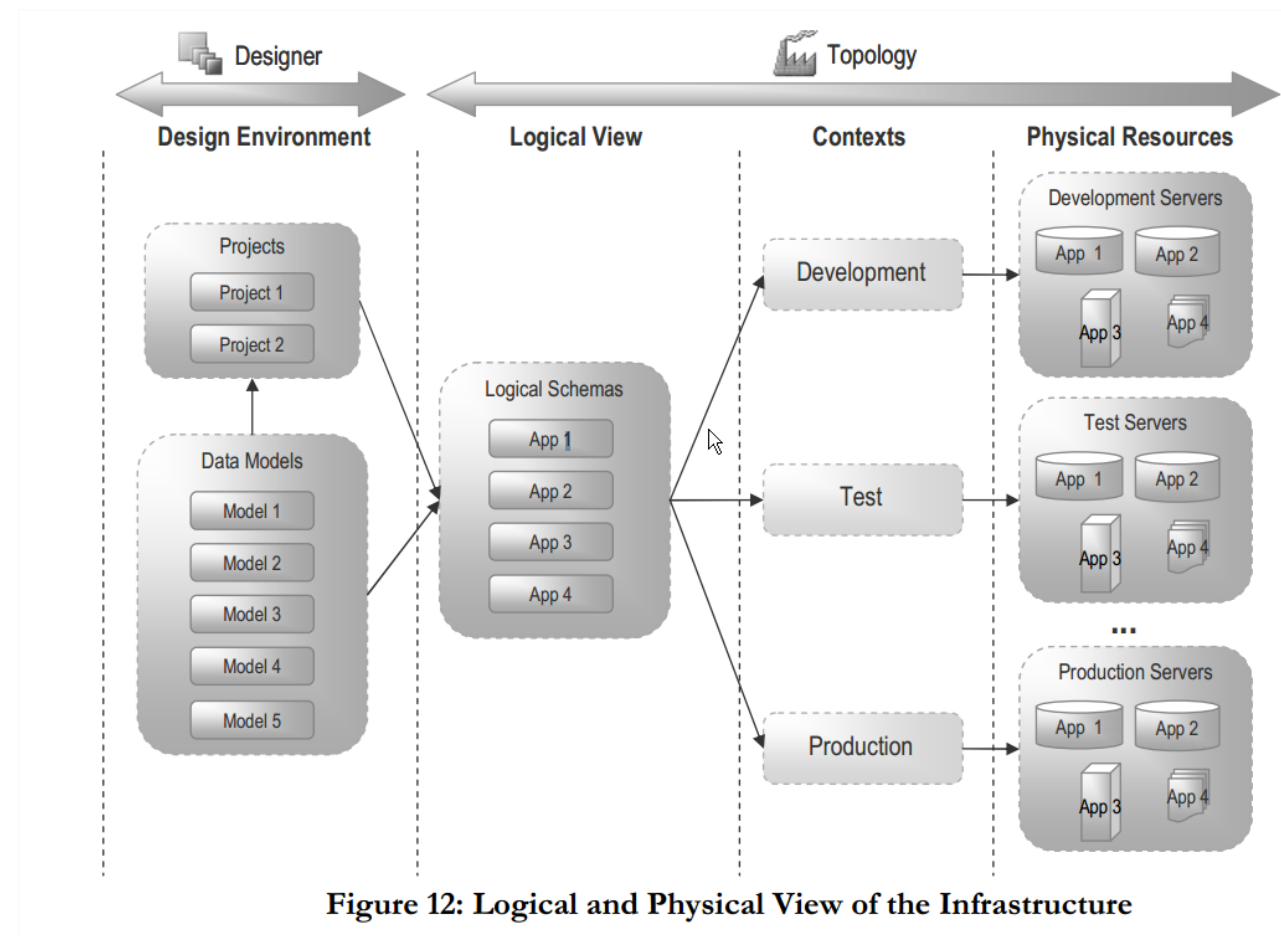
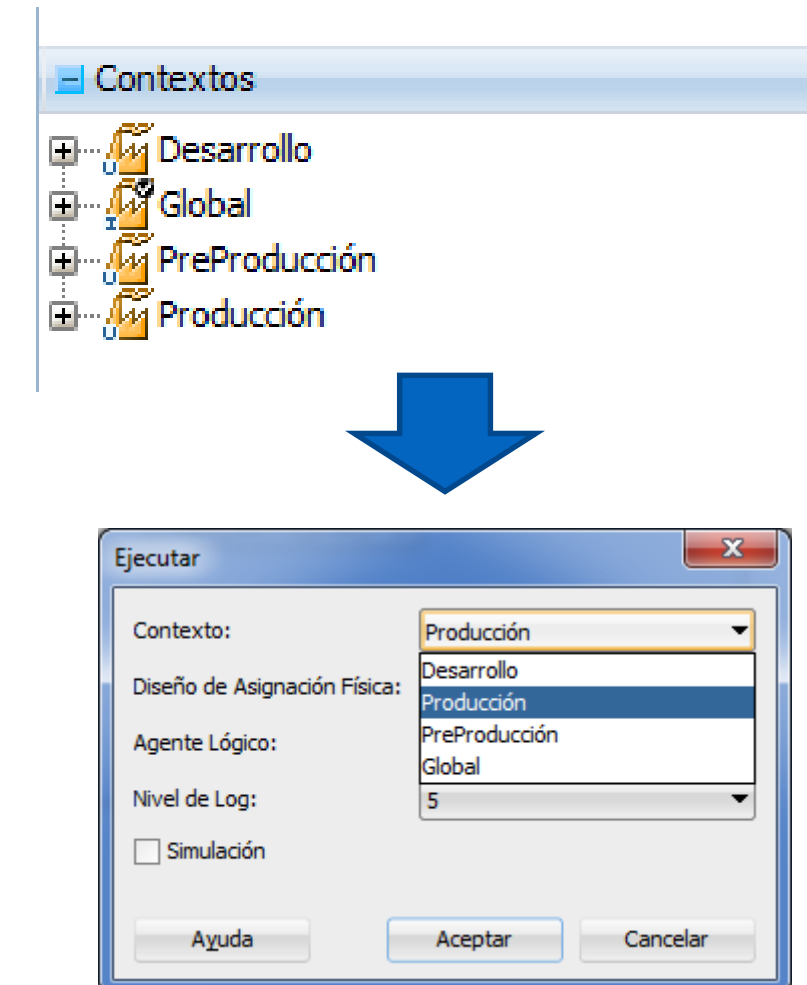


Figure 12: Logical and Physical View of the Infrastructure



# Data Management:

Gestión holística de los datos empresariales

## DATOS MAESTROS Y DE REFERENCIA



**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.

# Gestión de Datos Maestros y De Referencia



M

Los **Datos Maestros** se refieren a los datos acordados y compartidos en toda la empresa

Ejemplo: Datos de Clientes, Empleados, Productos, etc.



R

Los **Datos de Referencia** son un subconjunto de Datos Maestros que hacen referencia a los datos que definen el conjunto de valores permitidos que pueden ser utilizados por otros campos de datos.

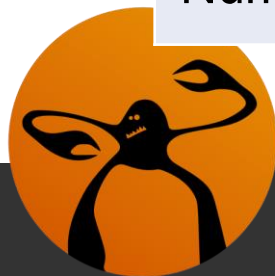
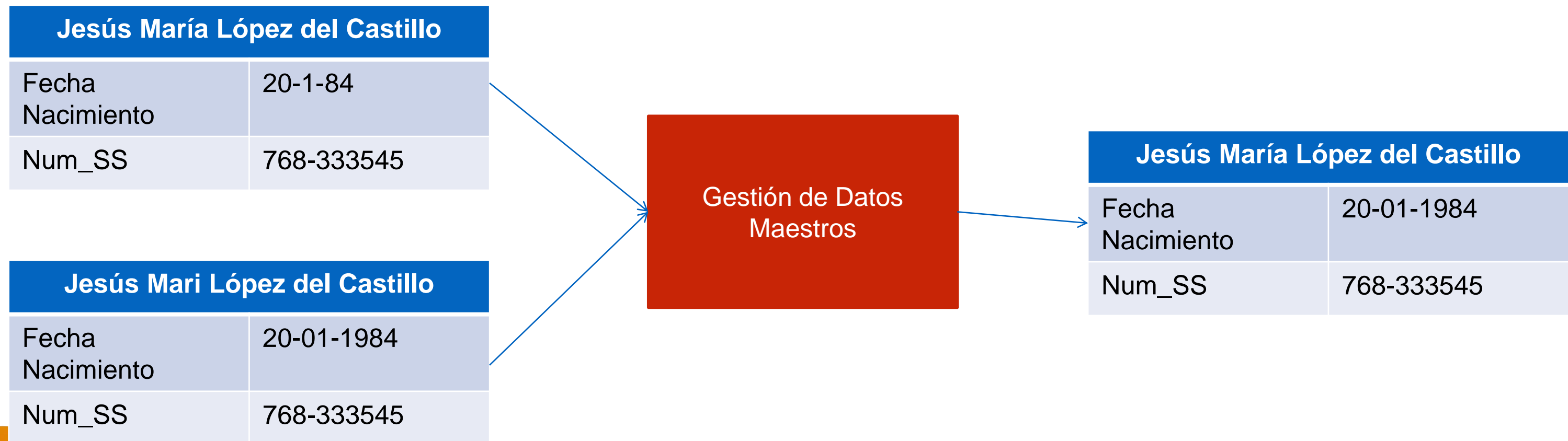
Ejemplo: Códigos de Países, Clasificación de Actividades Industriales, etc.





# Proceso de Gestión de Datos Maestros

- 1 Identificar registros con potenciales coincidencias
- 2 MDM aplica las reglas de negocio para combinar y fusionar los registros.
- 3 MDM crea el registro maestro con atributos de confianza



# Proceso de Gestión de Datos de Referencia

1

RDM determina registros para búsqueda

2

RDM proporciona tablas de referencia para las búsquedas

3

RDM actualiza los registros al estándar de negocio

Fuente 1

Jesús María López del Castillo	
Estado	FL
Ciudad	Orlando

Fuente 2

Ángel Ortega	
Estado	N.York
Ciudad	Albany

Tabla de Referencia de Datos (Estado)		
Valor Válido	Fuente 1	Fuente 2
Florida	FL	Florida
New York	NY	N.York

Fuente 1

Jesús María López del Castillo	
Estado	Florida
Ciudad	Orlando

Fuente 2

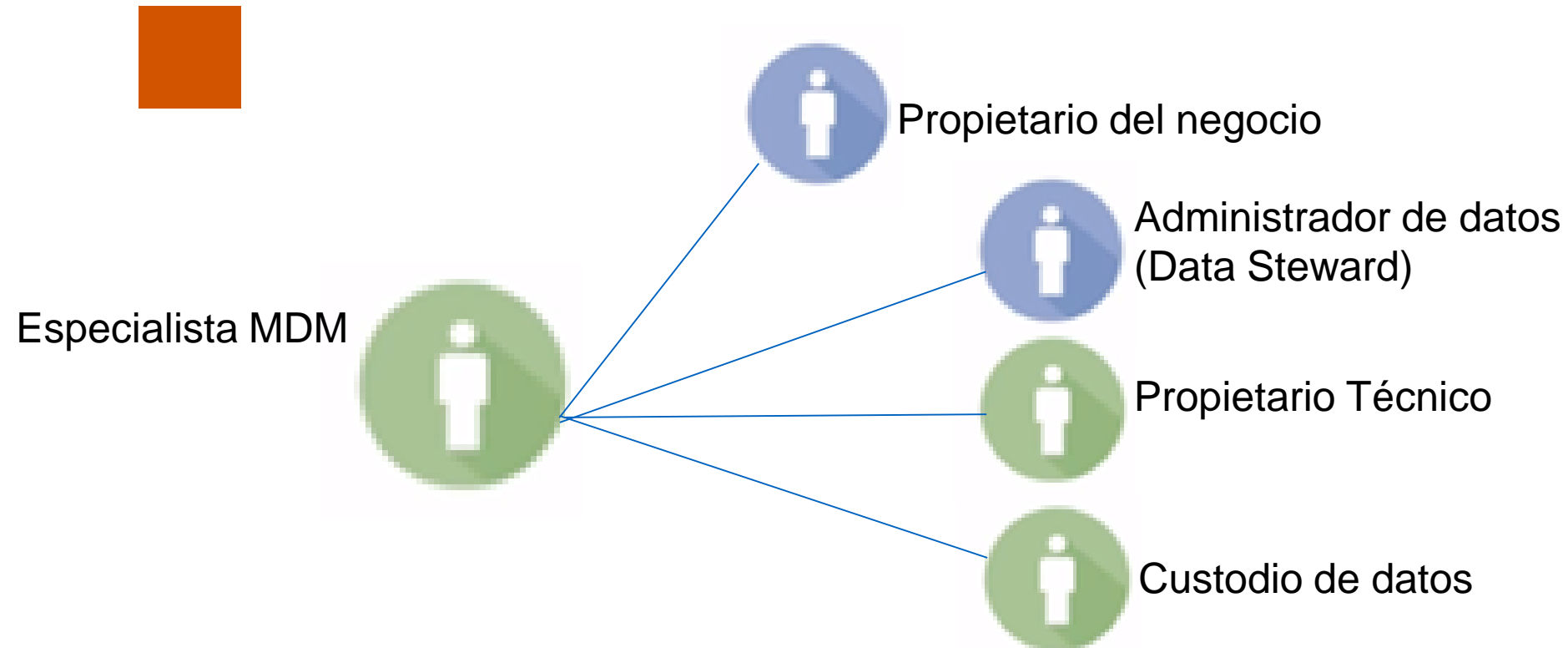
Ángel Ortega	
Estado	New York
Ciudad	Albany



# Roles de Gestión de Datos Maestros y De Referencia



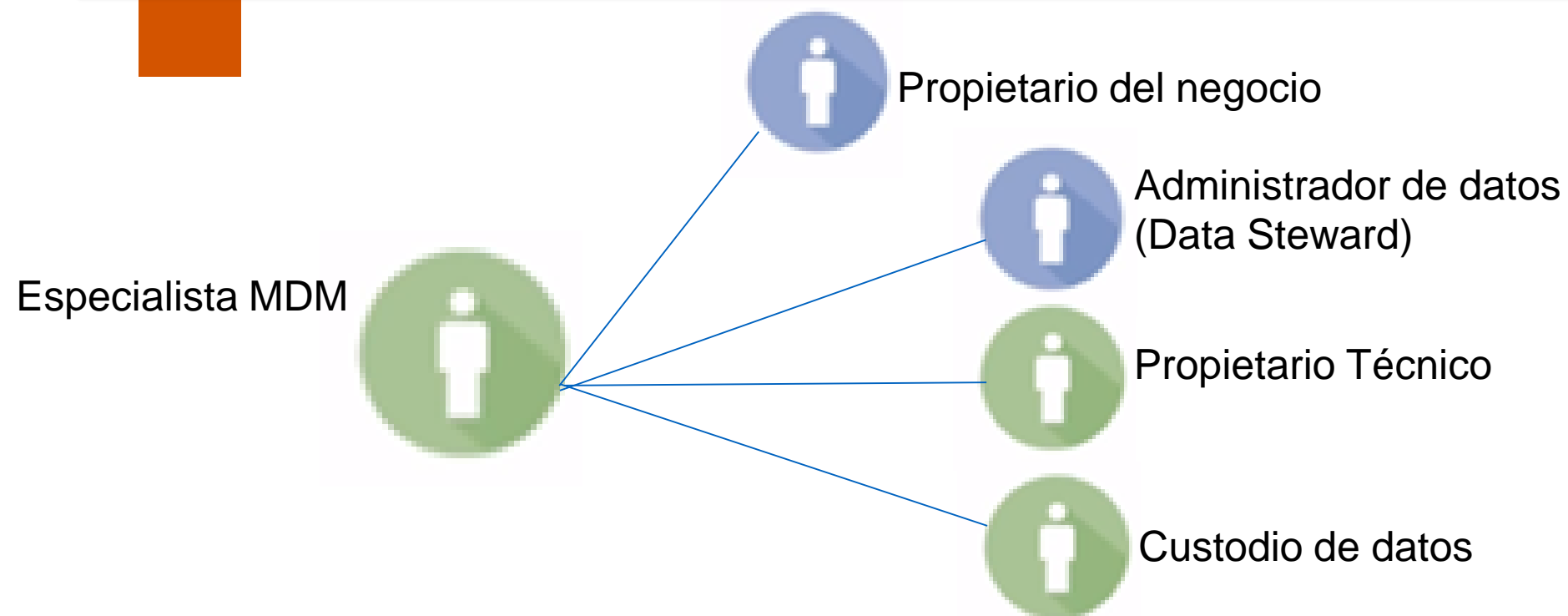
El **Especialista en MDM** es rol principal responsable de realizar las actividades asociadas con la gestión de Datos Maestros y De Referencia. Aunque es un rol específico a la gestión de Datos Maestros y De Referencia, Los especialistas en MDM trabajarán estrechamente con los propietarios de negocio, administradores de datos, propietarios técnicos y custodios de datos. Sus responsabilidades principales se refieren al diseño e implementación del producto MDM ( por ejemplo, definición de atributos en el MDM, creación documentos de mappings MDM, etc.)



# Roles de Gestión de Datos Maestros y De Referencia

 El **Especialista en MDM** es rol principal responsable de realizar las actividades asociadas con la gestión de Datos Maestros y De Referencia. Este es un rol específico a la gestión de Datos Maestros y De Referencia. Los especialistas en MDM trabajarán estrechamente con los propietarios de negocio, administradores de datos, propietarios técnicos y custodios de datos,. Sus responsabilidades principales se refieren al diseño e implementación de aplicaciones de gestión Datos Maestros ( por ejemplo, definición de atributos en el MDM, creación documentos de mappings MDM, etc.)

## PERSONAS



*La definición del Rol, Especialista MDM, y sus responsabilidades cubren la perspectiva de personas dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management*



Rol de Negocio



Rol Técnico



# Herramientas de Datos Maestros y De Referencia

## Las características principales de las herramientas de Gestión de Datos Maestros y De Referencia.

- Listo para utilizar en múltiples dominios como Cliente, Producto, Localización y Cuentas.
- Gestión eficiente de la relación entre dominios, por ejemplo, cliente a productos.
- Categorización, agrupación y jerárquica de entidades de datos maestros.
- Potente algoritmo de duplicación de datos para identificar y eliminar registros duplicados.
- Capacidad de configurar fácilmente los elementos de datos críticos necesarios para la coincidencia de acuerdo con los requisitos de la organización.
- Capacidad de configurar reglas de negocios para mantener la información actualizada y más reciente precisa, lo que resulta en la creación de un registro de datos único.

## HERRAMIENTAS DE DATOS MAESTROS y DE REFERENCIA

GESTIÓN DE  
DATOS MAESTROS

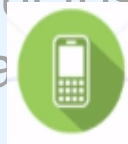
GESTIÓN DE  
DATOS DE  
REFERENCIA



# Herramientas de Datos Maestros y De Referencia

Las características principales de las herramientas de Gestión de Datos Maestros y De Referencia.

- Listo para utilizar en múltiples dominios como Cliente, Producto, Localización y Cuentas.
- Gestión eficiente de la relación entre dominios, por ejemplo, cliente a productos.
- Categorización, agrupación y jerárquica de entidades de datos maestros.
- Potente algoritmo de duplicación de datos para identificar y eliminar registros duplicados.
- Capacidad de configurar fácilmente los elementos de datos críticos necesarios para la coincidencia de acuerdo con los requisitos de la organización.
- Capacidad de configurar reglas de negocios para mantener la información actualizada y más reciente precisa, lo que resulta en la creación de un registro de datos único.



## TECNOLOGÍA

HERRAMIENTAS DE  
DATOS MAESTROS y DE REFERENCIA

GESTIÓN DE  
DATOS MAESTROS

GESTIÓN DE  
DATOS DE  
REFERENCIA

*Las Herramientas Tecnológicas que dan soporte a la gestión de maestros y de referencia hacen referencia a la **perspectiva de tecnología** dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management*





# Resumen – Puntos Clave

Los **Datos Maestros** se refieren a los datos acordados y compartidos en toda la organización

Los **Datos de Referencia** son un subconjunto de Datos Maestros que hacen referencia a los datos que definen el conjunto de valores permitidos que pueden ser utilizados por otros campos de datos.

La gestión de datos maestros y de referencia ayuda al **proceso de estandarización de datos** en toda la organización

Las características principales de las **herramientas de MDM** están relacionadas con la clasificación y de duplicación de datos.

El Rol clave en la gestión de datos maestros y de referencia es un **especialista en MDM**

Ejemplos típicos de **Datos maestros** : **Cliente, Producto y Empleado**. Ejemplos típicos de **Datos de referencia** : **Género, Ciudad, Provincia, Código Postal y Estado Civil**



# Data Management:

Gestión holística de los datos empresariales

## SEGURIDAD DE DATOS



**Nerea Sevilla**, Responsable de Proyectos de Analítica en Lanbide, Servicio Vasco de Empleo.



# Data Management:

## Gestión holística de los datos empresariales

Smarter With Gartner

Subscribe



Gartner

August 24, 2018 | Contributor: Laurence Goasduff

### CISOs Must Master Data Governance for GDPR Compliance

Security



#### The new EU regulation is a wake-up call to draft new data security strategies.

The European Union's [General Data Protection Regulation \(GDPR\)](#), which comes with potentially high fines for noncompliance, is forcing chief information security officers (CISOs) at organizations of all sizes to rethink how they manage data privacy. Yet many still don't have a data security governance strategy.

"Although GDPR guidelines have been in effect since 25 May 2018, it's clear that many organizations lack such a strategy or the tools needed to effectively protect sensitive data and maintain privacy and protection," says [Deborah Kish](#), principal research analyst at Gartner.

"GDPR is a wake-up call for CISOs to draft new data security strategies"



CISO: Oficial Principal de Seguridad de la información / Chief Information Security Officer

# Gestión de la Seguridad de datos

La **Seguridad de los Datos** ( también llamada Protección de Datos o Privacidad de datos), hace referencia a los procesos, políticas y la tecnología necesarias para proteger la información confidencial de accesos y usos no autorizados, interna y externamente.

## Ejemplos de información confidencial



Número de seguridad social

fecha de nacimiento

origen étnico

número de tarjeta de crédito

Plan de ventas



# Información Sensible vs. Información No Sensible

## Información No-Sensible

**Información pública** es la información que ya es de dominio o conocimiento público (por ejemplo, archivos de registro de delincuentes sexuales y censo de votantes).

**Información rutinaria de negocio**, es la información del negocio que no está sujeta a una ninguna protección especial y que puede compartirse de forma rutinaria con cualquier persona dentro o fuera de la empresa.

## Información Sensible

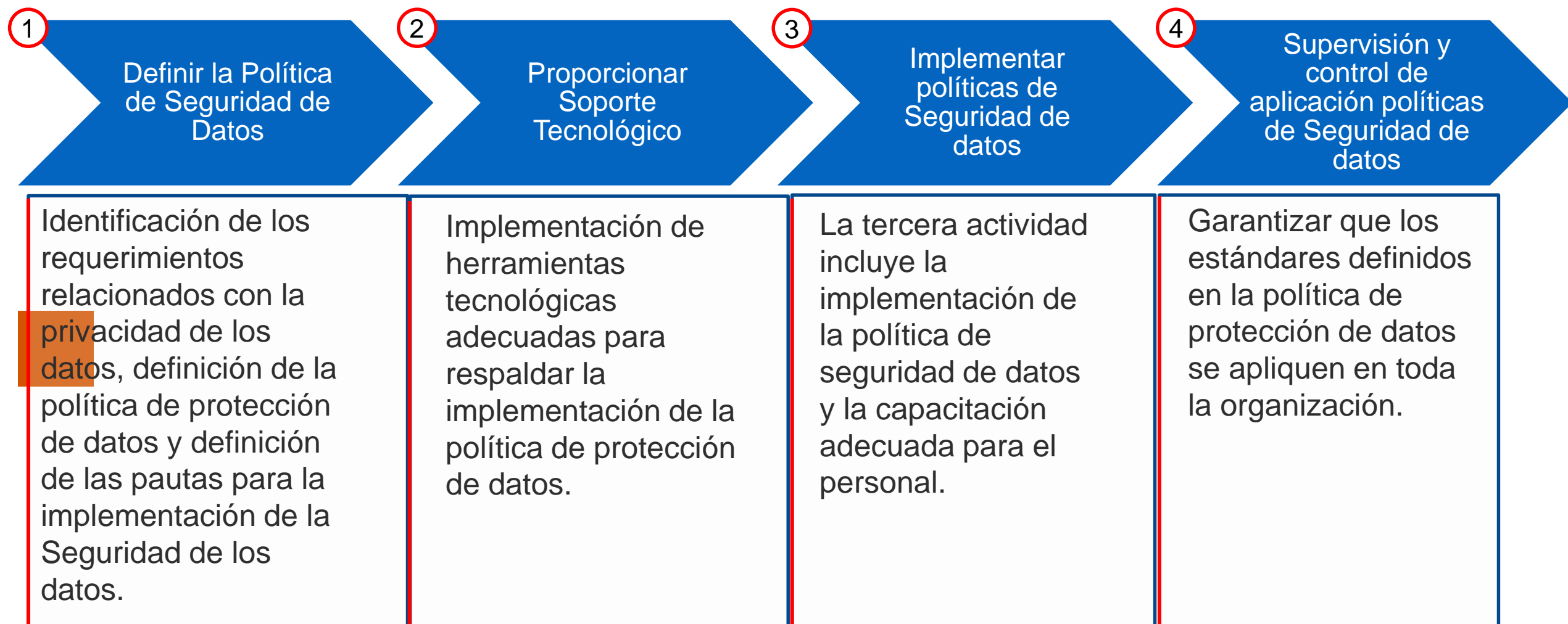
**Información personal y privada** se refiere a la información que pertenece a una persona privada, pero la persona puede decidir compartirla con otras personas por razones personales o de negocio (por ejemplo, número de seguro social).

**Información Confidencial del negocio**, es la información cuya divulgación puede afectar a la empresa (por ejemplo, planes de ventas y marketing).

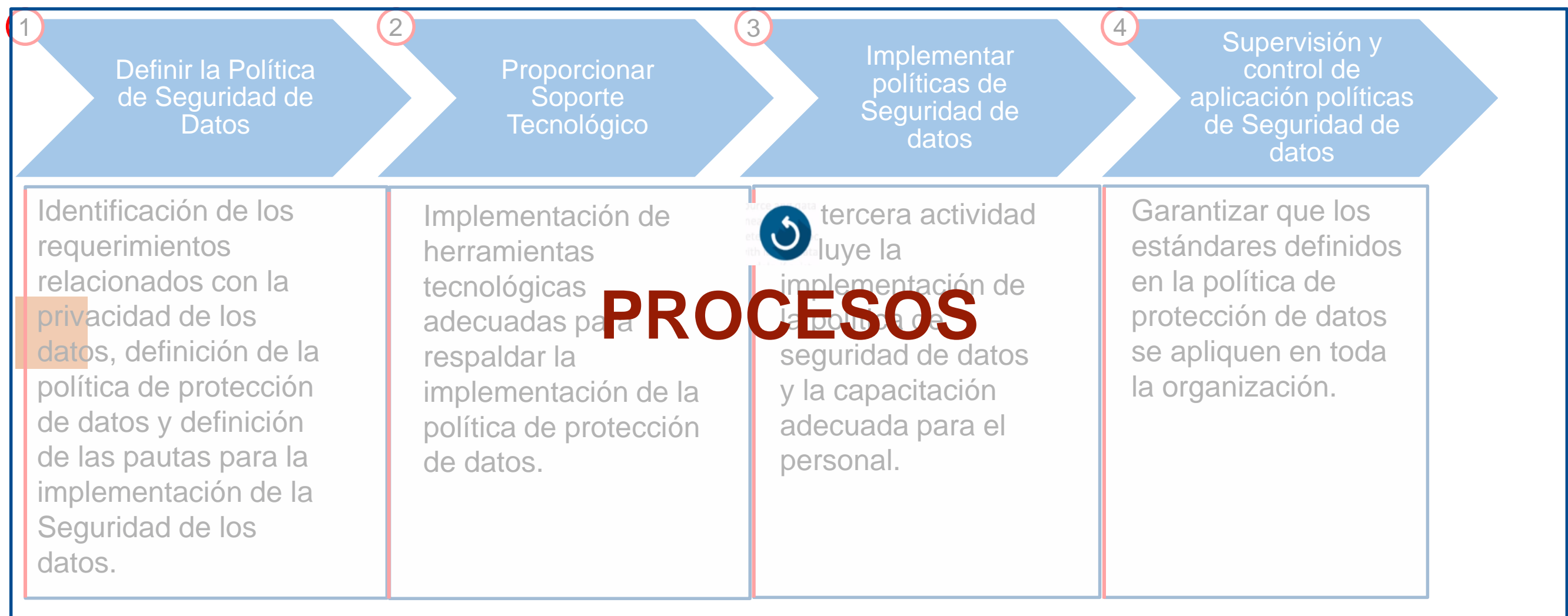
**Información Clasificada** generalmente es la información que está sujeta a la clasificación de Regulaciones Especiales de Seguridad impuestas por muchos gobiernos nacionales.



# Proceso de Seguridad de datos



# Proceso de Seguridad de datos



*Las actividades del proceso de seguridad de datos cubren la perspectiva de procesos dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management*

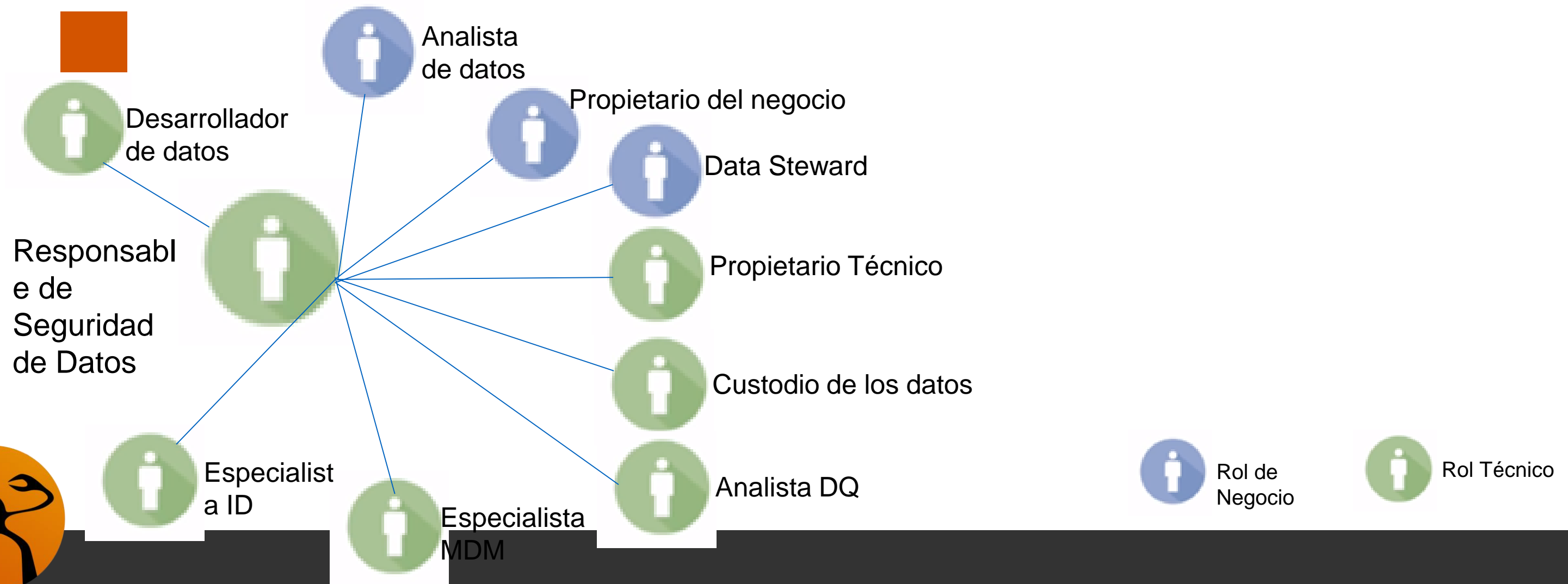




# Roles de Seguridad de datos



El **Responsable de Seguridad de datos** es el rol clave responsable de implementar políticas de protección de datos. Aunque esta función es específica de la seguridad de datos, el responsable de privacidad de datos trabajará en estrecha colaboración con todas las demás funciones de administración de datos. Las responsabilidades clave incluyen: 1) Desarrollar e implementar la política de seguridad de datos, 2) Proporcionar información y orientación sobre el procesamiento de todos los datos personales, 3) Desarrollar una Guía con “Mejores Practicas” para el personal, 4) Impartir capacitación al personal y 5) Procesar, coordinar y responder a todas las solicitudes de información.

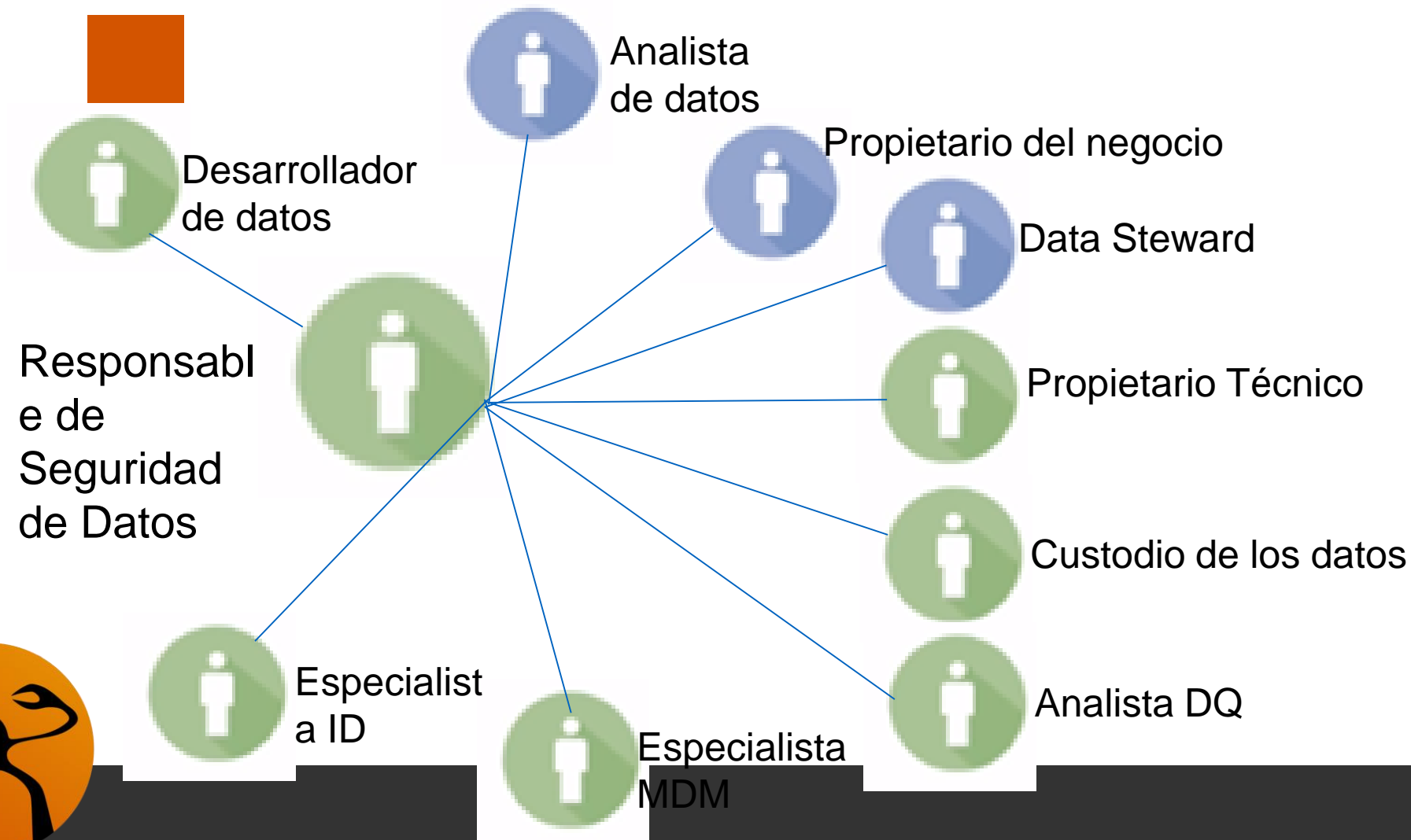


# Roles de Seguridad de datos



El **Responsable de Seguridad de datos** es el rol clave responsable de implementar políticas de protección de datos. Aunque esta función es específica de la seguridad de datos, el responsable de privacidad de datos trabajará en estrecha colaboración con todas las demás funciones de administración de datos. Las responsabilidades clave incluyen: 1) Desarrollar e implementar la política de seguridad de datos, 2) Proporcionar información y orientación sobre el procesamiento de todos los datos personales, 3) Desarrollar una Guía con “Mejores Practicas” para el personal, 4) Impartir capacitación al personal y 5) Procesar, coordinar y responder a todas las solicitudes de información.

## PERSONAS



*La definición del Responsable de Seguridad de datos y sus responsabilidades cubren la perspectiva de personas dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management*




Rol de Negocio



Rol Técnico



# Herramientas Tecnológicas para Seguridad de datos



Protección del correo electrónico.

- Filtrado de mensajes entrantes, filtrado y borrado de spam automáticamente, análisis de archivos adjuntos y también cifrado de comunicaciones de business-to-Business para mejorar la seguridad.

Antivirus .

- para los empleados que tienen acceso a Internet, se deberá tener un programa antivirus que se ejecute en todas los ordenadores de la red.

Firewall.

- Un sistema hardware de firewall aumenta la seguridad de todas en la red de ordenadores en la organización.

Wi-Fi Encriptado.

- Si existe una configuración de red Wi-Fi en la empresa, la red deberá ser encriptada con acceso restringido a personal autorizado con mediante código de acceso..

Almacenamiento de datos en la nube.

- Guardar y trabajar con documentos en la nube, hay que habilitar la herramienta de seguridad en la nube para facilitar el acceso de múltiples usuarios a documentos y archivos desde un dispositivo móvil o desde una ubicación remota.

Acceso seguro a la red.

- Protegiendo el acceso a Internet en la organización porque es importante proteger los servidores y sistemas a los ataques de hackers






# Herramientas Tecnológicas para Seguridad de datos





*Las Herramientas Tecnológicas que dan soporte a la gestión de seguridad de datos hacen referencia a la **perspectiva de tecnología** dentro de las perspectivas de personas, procesos y tecnología existentes en el Data Management*


# Resumen – Puntos Clave




**Seguridad de los datos** ( también denominada protección de datos o privacidad de datos), hace referencia procesos, políticas y tecnología para proteger la información confidencial de accesos y usos no autorizados, interna y externamente



La **Información No Confidencial** se refiere a la información pública y la información rutinaria del negocio.



La **Información Confidencial** se refiere a información personal y privada, información confidencial del negocio e información clasificada.



Los Componentes Tecnológicos necesarios para respaldar la privacidad de los datos son: **protección de correo electrónico, antivirus, firewall, sistema de almacenamiento en la nube y acceso a la seguridad de la red.**



El **RoI** clave en la seguridad de datos es el **Responsable de Seguridad de datos.**

