

Data Management Fundamentals

Módulo 7 - Data Integration & Interoperability

Alberto Turégano

GOLD SPONSORS



SILVER SPONSORS



Entidades académicas



Universidad de
Castilla-La Mancha



Alberto Turégano



- Ingeniero informático UPM y Máster en tecnologías Big Data y Analytics
- Toda una vida dedicada al mundo de dato
- Asesor *freelance* ayudando a implantar la cultura del dato en las pymes
- Director del Master en Big Data Analytics de EOI
- Profesor Data Storytelling IEB, The Valley, etc.
- Profesor de Gestión y Gobierno del Dato en IE University, UFV
- Otras aficiones: Instructor PADI, apneista, triatleta, cartomago...

Definición



- DI&I describe los procesos relativos al **movimiento** y la **consolidación de datos** dentro y entre almacenes de datos, aplicaciones y organizaciones.

Definición



- DI&I describe los procesos relativos al movimiento y la consolidación de datos dentro y entre almacenes de datos, aplicaciones y organizaciones.
- La integración consolida los datos en formas consistentes, ya sean físicas o virtuales.

Definición



- DI&I describe los procesos relativos al movimiento y la consolidación de datos dentro y entre almacenes de datos, aplicaciones y organizaciones.
- La integración consolida los datos en formas consistentes, ya sean físicas o virtuales.
- La Interoperabilidad de Datos es la capacidad para que múltiples sistemas se comuniquen.

Las soluciones de DII permiten funciones básicas:



- Migración y conversión de datos
- Consolidación de datos en nodos (hubs) o data marts
- Integración de paquetes de proveedores en la cartera de aplicaciones de una organización
- Intercambio de datos entre aplicaciones y entre organizaciones
- Distribución de datos entre almacenes y centros de datos
- Archivado de datos
- Gestión de interfaces de datos
- Obtención e ingesta de datos externos
- Integración de datos estructurados y no estructurados
- Proporcionar inteligencia operativa y apoyo a la toma de decisiones de gestión

Áreas que interactúan con DII



Áreas que interactúan con DII: Data Governance



Áreas que interactúan con DII: Data Governance



Áreas que interactúan con DII: Data Governance



Áreas que interactúan con DII: Data Governance



Áreas que interactúan con DII: Data Governance



Áreas que interactúan con DII: Data Governance



Áreas que interactúan con DII: Data Governance



Motivadores de negocio

Data Integration and Interoperability

Definition: Data Integration is the movement and consolidation of data within and between data stores, applications and organizations. Data Interoperability is the ability for multiple systems to communicate.

Goals:

1. Make data available from disparate sources in the format and timeframe needed.
2. Lower cost and complexity of managing solutions by developing shared models and interfaces.
3. Identify meaningful events and automatically trigger alerts and actions.
4. Support business intelligence, analytics, master data management, and operational efficiency efforts.

- Eficiencia en la Gestión de Datos.



(P) Planning, (C) Control, (D) Development, (O) Operations

Motivadores de negocio

Data Integration and Interoperability

Definition: Data Integration is the movement and consolidation of data within and between data stores, applications and organizations. Data Interoperability is the ability for multiple systems to communicate.

Goals:

1. Make data available from disparate sources in the format and timeframe needed.
2. Lower cost and complexity of managing solutions by developing shared models and interfaces.
3. Identify meaningful events and automatically trigger alerts and actions.
4. Support business intelligence, analytics, master data management, and operational efficiency efforts.

- Eficiencia en la Gestión de Datos.
- Integración de Aplicaciones.



(P) Planning, (C) Control, (D) Development, (O) Operations

Motivadores de negocio

Data Integration and Interoperability

Definition: Data Integration is the movement and consolidation of data within and between data stores, applications and organizations. Data Interoperability is the ability for multiple systems to communicate.

Goals:

1. Make data available from disparate sources in the format and timeframe needed.
2. Lower cost and complexity of managing solutions by developing shared models and interfaces.
3. Identify meaningful events and automatically trigger alerts and actions.
4. Support business intelligence, analytics, master data management, and operational efficiency efforts.



(P) Planning, (C) Control, (D) Development, (O) Operations

- Eficiencia en la Gestión de Datos.
- Integración de Aplicaciones.
- Reducir el coste de soporte.

Metas y principios

Metas



Metas y principios

Metas



- Hacer que los datos estén **disponibles** en el formato y el plazo que necesitan los consumidores de datos (tanto humanos como sistemas)

Metas y principios

Metas



- Hacer que los datos estén **disponibles** en el formato y el plazo que necesitan los consumidores de datos (tanto humanos como sistemas)
- **Consolidar** los datos física y virtualmente en centros de datos.

Metas y principios

Metas



- Hacer que los datos estén **disponibles** en el formato y el plazo que necesitan los consumidores de datos (tanto humanos como sistemas)
- **Consolidar** los datos física y virtualmente en centros de datos.
- **Reducir el coste** y la complejidad de la gestión de las soluciones mediante el desarrollo de modelos e interfaces compartidos.

Metas y principios

Metas



- Hacer que los datos estén **disponibles** en el formato y el plazo que necesitan los consumidores de datos (tanto humanos como sistemas)
- **Consolidar** los datos física y virtualmente en centros de datos.
- **Reducir el coste** y la complejidad de la gestión de las soluciones mediante el desarrollo de modelos e interfaces compartidos.
- **Identificar eventos significativos** (oportunidades y amenazas) y activar automáticamente alertas y acciones.

Metas y principios

Metas



- Hacer que los datos estén **disponibles** en el formato y el plazo que necesitan los consumidores de datos (tanto humanos como sistemas)
- **Consolidar** los datos física y virtualmente en centros de datos.
- **Reducir el coste** y la complejidad de la gestión de las soluciones mediante el desarrollo de modelos e interfaces compartidos.
- **Identificar eventos significativos** (oportunidades y amenazas) y activar automáticamente alertas y acciones.
- **Apoyar el Business Intelligence**, la analítica, la gestión de datos maestros y la eficiencia operativa.

Metas y principios

Principios



Metas y principios

Principios



- Adoptar **una perspectiva empresarial en el diseño** para garantizar la extensibilidad futura, pero impleméntela a través de la entrega iterativa e incremental

Metas y principios

Principios



- Adoptar **una perspectiva empresarial en el diseño** para garantizar la extensibilidad futura, pero impleméntela a través de la entrega iterativa e incremental
- **Equilibrar las necesidades de datos locales con las necesidades de datos empresariales**, incluidos el soporte y el mantenimiento.

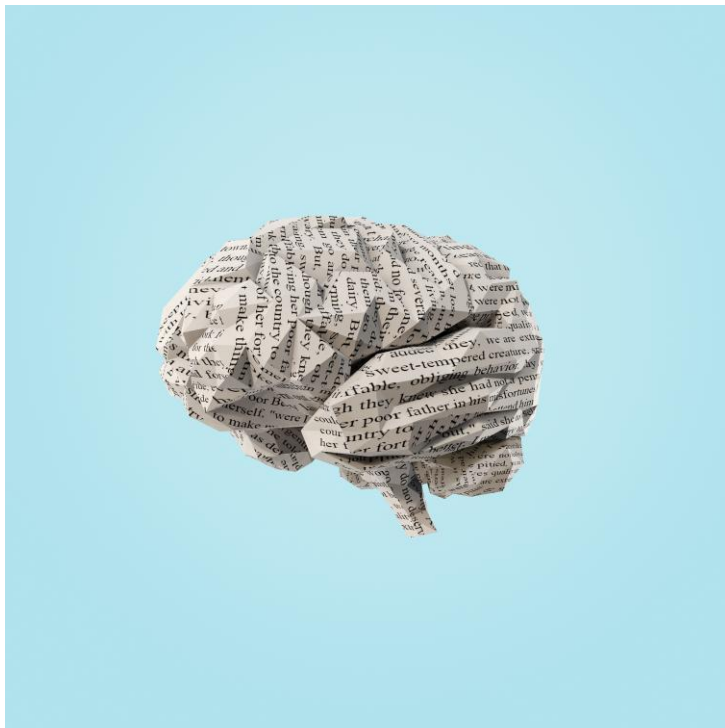
Metas y principios

Principios



- Adoptar **una perspectiva empresarial en el diseño** para garantizar la extensibilidad futura, pero impleméntela a través de la entrega iterativa e incremental
- **Equilibrar las necesidades de datos locales con las necesidades de datos empresariales**, incluidos el soporte y el mantenimiento.
- Garantizar la responsabilidad empresarial para el diseño y la actividad de DII. **Los expertos en negocio deben participar en el diseño y la modificación de las reglas de transformación de datos**, tanto persistentes como virtuales.

Conceptos esenciales

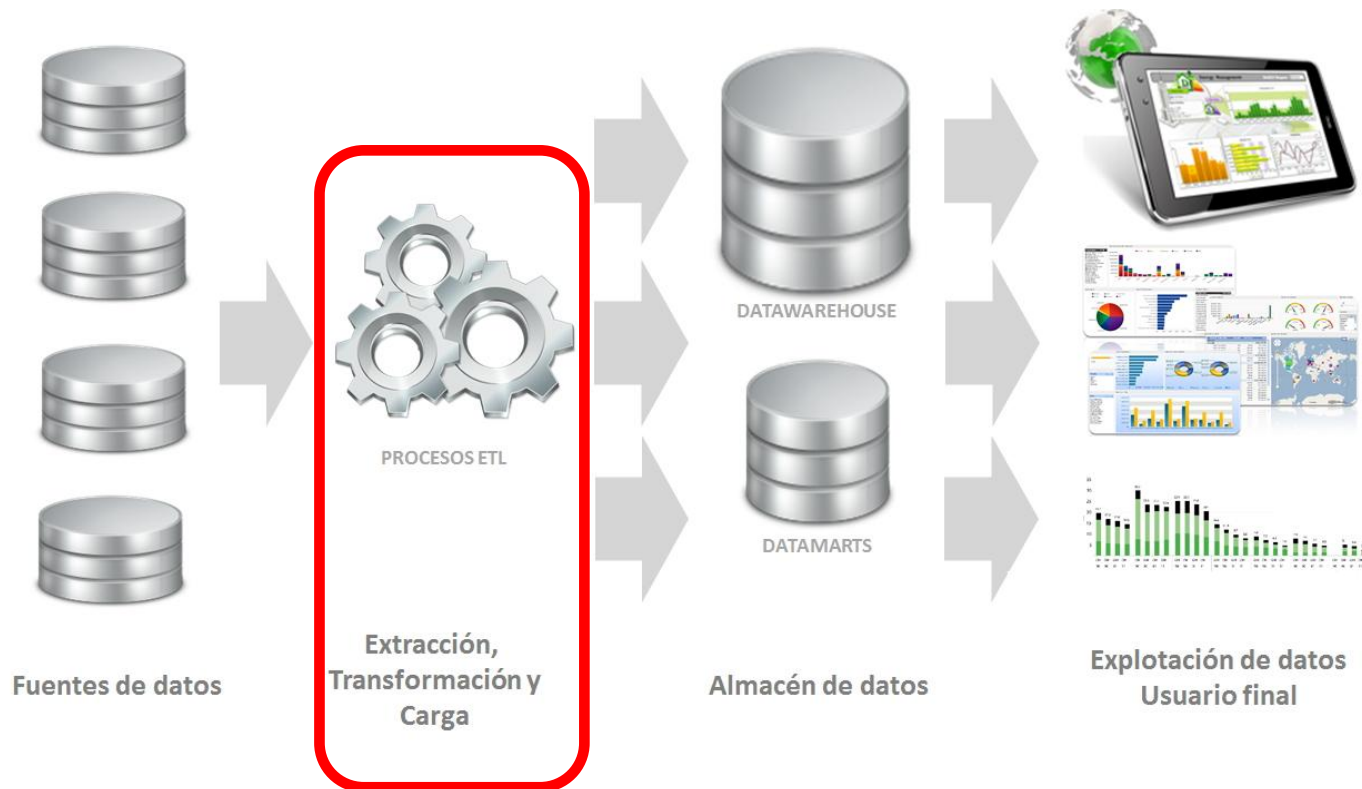


- Extraer, Transformar y Cargar

Arquitectura clásica BI



Arquitectura clásica BI ⚠



Extracción

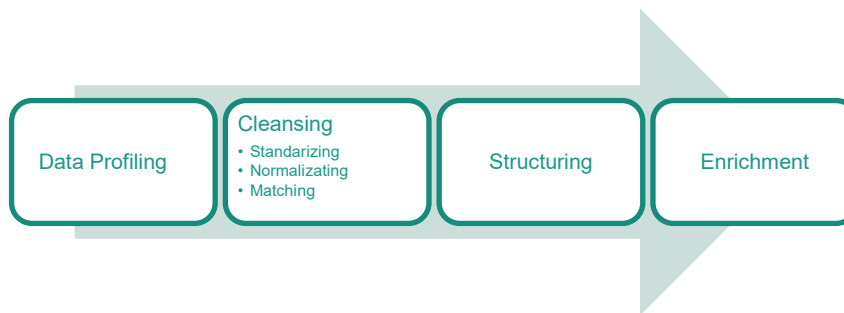


- Se puede realizar la extracción mediante rutinas ad-hoc o utilizando herramientas comerciales
- Problemas al acceder a las fuentes:
 - Diferentes BB.DD (Host, relacionales, ficheros, etc.)
 - Diferentes Plataformas (Host, UNIX, PCs, etc.)
 - Diferentes protocolos (TCP/IP, IPX, Propietarios, etc.)
 - Diferentes juegos de caracteres (EBCDIC, ASCII, etc.)
 - Diferentes tipos de datos (Fechas, horas, numéricos, booleanos, etc.)
- Diferentes estrategias de carga (todo o lo necesario para el modelo de negocio) en la zona de staging

Transformación



- Los sistemas operacionales tienen datos “sucios” por diferentes motivos que hay que limpiar:
 - Valores por defecto (ej: 999999999)
 - Ausencia de valor
 - Campos con diferentes utilidades
 - Selección del primer valor en una lista desplegable
 - Etc.
- La transformación de datos se divide (básicamente) a su vez en cuatro etapas:

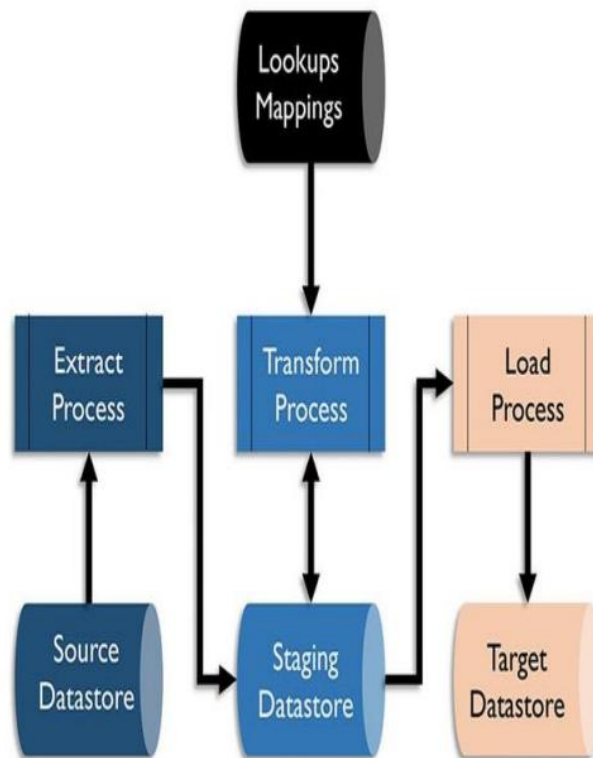


Carga

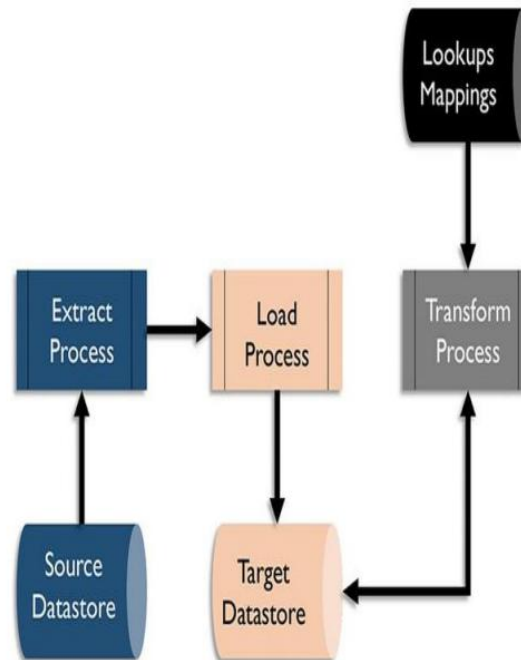
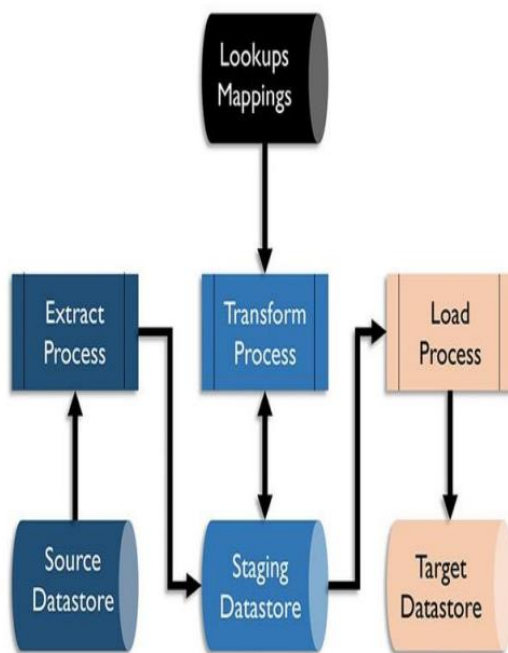


- En esta fase de integración de la información en el modelo final explotable por los usuarios se verifica que la información es correcta en base a modelos de calidad con métricas establecidas por los mismos usuarios de negocio.

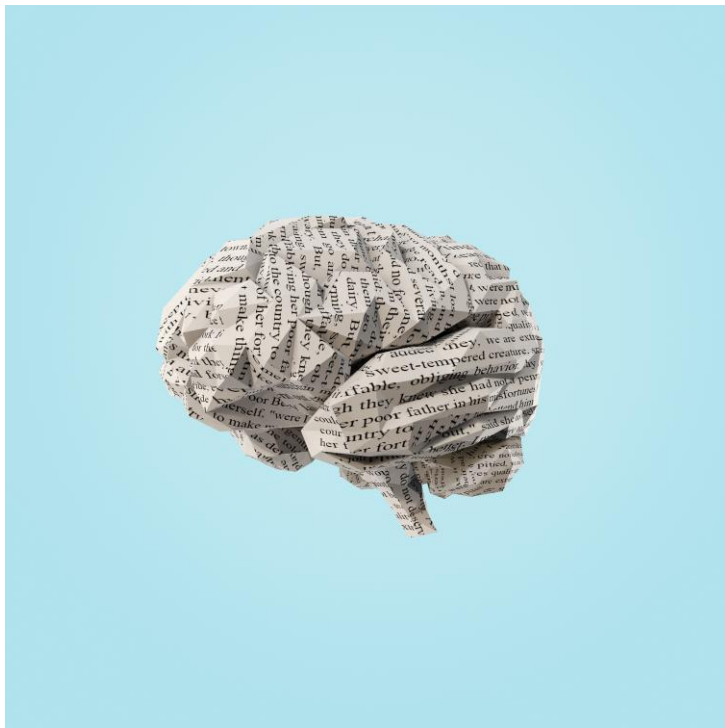
ETL



ETL vs ELT

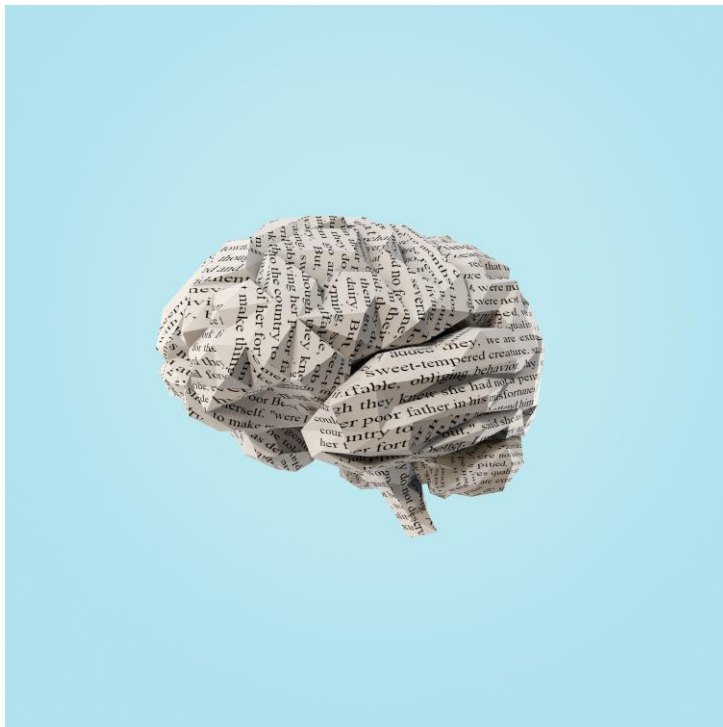


Conceptos esenciales



- Extraer, Transformar y Cargar

Conceptos esenciales



- Extraer, Transformar y Cargar
- Latencia

Latency

La latencia es la diferencia de tiempo entre el momento en que se generan los datos en el sistema de origen y el momento en que los datos están disponibles para su uso en el sistema de destino.

- Los distintos enfoques del procesamiento de datos dan lugar a diferentes grados de latencia de los datos.
- La latencia puede ser alta (por lotes) o baja (basada en eventos) hasta muy baja (sincrónica en tiempo real).

Tipología de latencias

Tipología de latencias

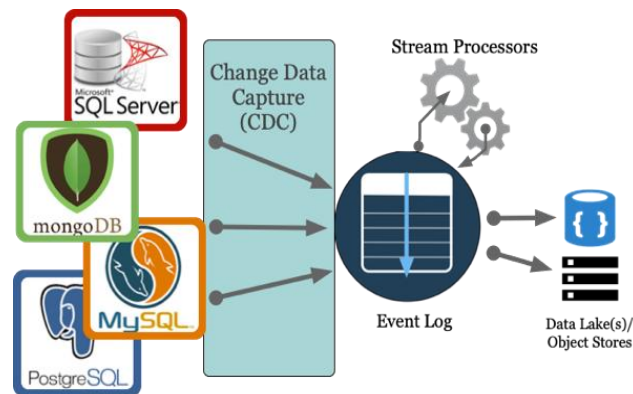
- Batch (por lotes)

The ETL Process



Tipología de latencias

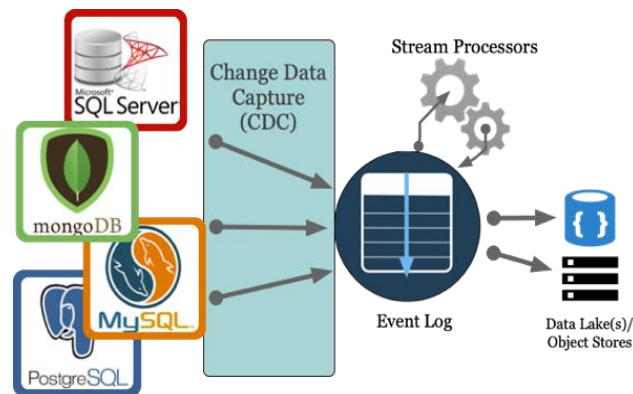
- Batch (por lotes)
- CDC (Change Data Capture)



ID_Cliente	Nombre	Apellido	Email	Fecha_Creacion	Fecha_Actualizacion	Estado
1	Juan	PÃ@rez	juan.perez@mail.com	15/01/2023	15/01/2023	Activo
2	Ana	GÃ³mez	ana.gomez@mail.com	20/02/2023	01/06/2023	Activo
3	Luis	MartÃnez	luis.martinez@mail.com	05/03/2023	10/05/2023	Inactivo

Tipología de latencias

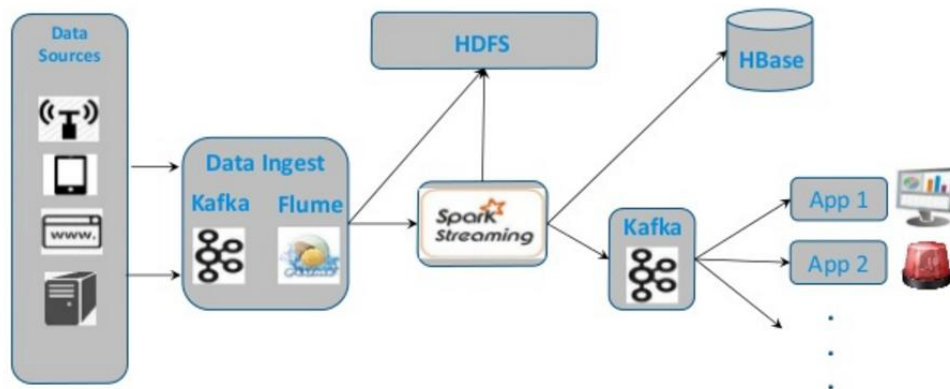
- Batch (por lotes)
- CDC (Change Data Capture)



Timestamp	Operation	Table	Key	Old Value	New Value
2024-06-14 10:15:01	INSERT	Orders	123	-	{OrderID: 123, CustomerID: 456, Amount: 100.00}
2024-06-14 10:16:45	UPDATE	Orders	123	{Amount: 100.00}	{Amount: 150.00}
2024-06-14 10:18:30	DELETE	Orders	124	{OrderID: 124, CustomerID: 789, Amount: 75.00}	-

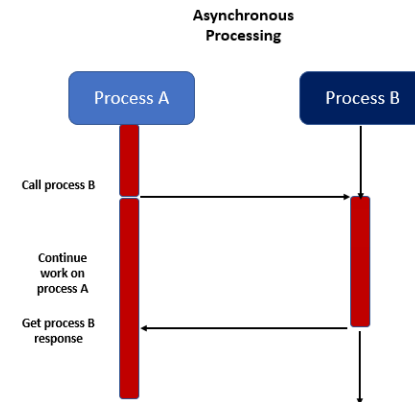
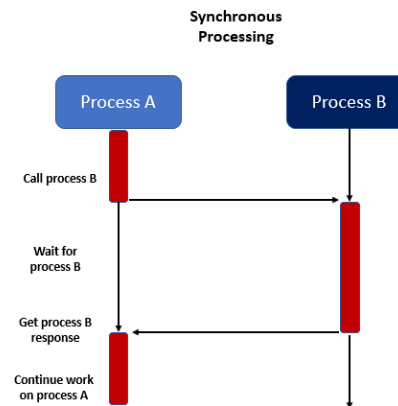
Tipología de latencias

- Batch (por lotes)
- CDC (Change Data Capture)
- **Event Driven (Near Real Time)**



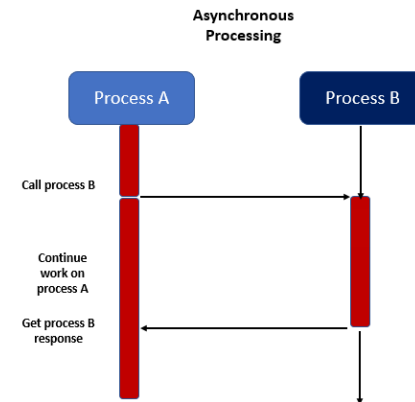
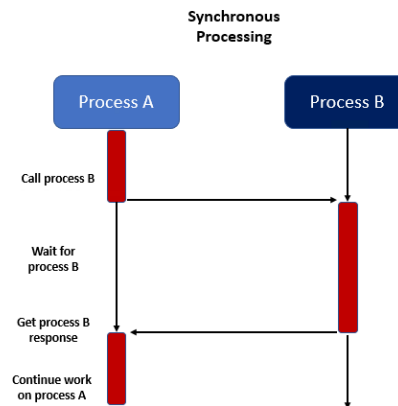
Tipología de latencias

- Batch (por lotes)
- CDC (Change Data Capture)
- Event Driven (Near Real Time)
- **Asynchronous**



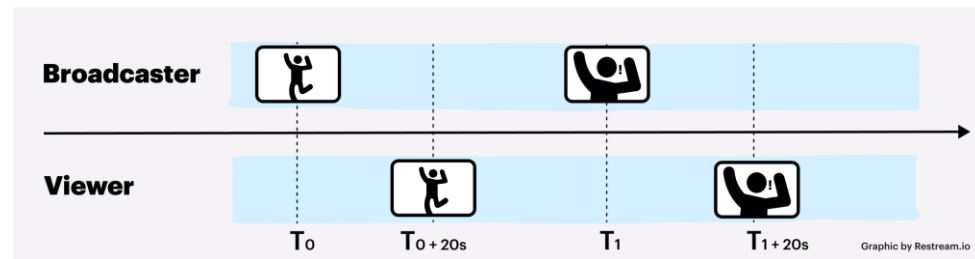
Tipología de latencias

- Batch (por lotes)
- CDC (Change Data Capture)
- Event Driven (Near Real Time)
- Asynchronous
- Real Time, Synchronous



Tipología de latencias

- Batch (por lotes)
- CDC (Change Data Capture)
- Event Driven (Near Real Time)
- Asynchronous
- Real Time, Synchronous
- Low Latency (Streaming)



Tipología de latencias

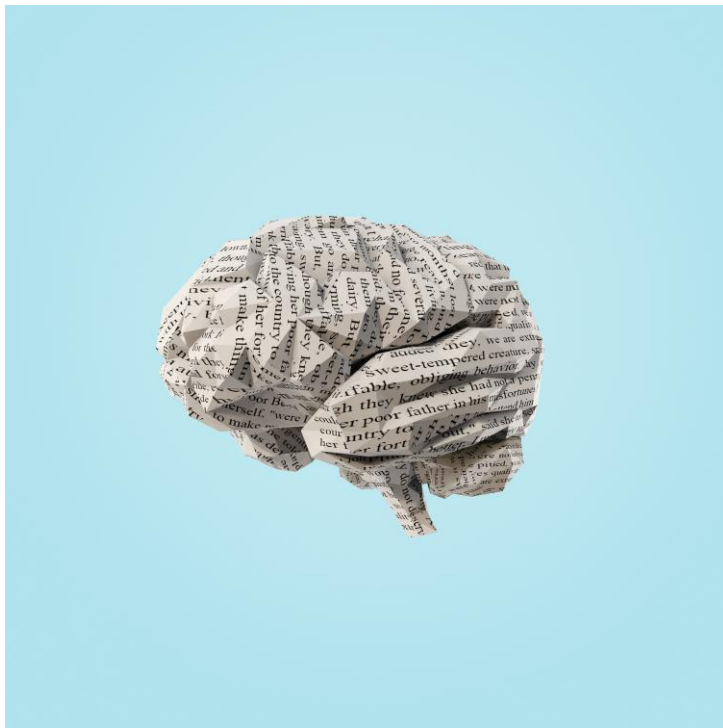
Tipología de Latencia	Descripción	Ejemplo	Tiempos de Actualización
Batch (por lotes)	Procesamiento de datos en grupos o lotes a intervalos regulares.	Generación de reportes financieros diarios.	Minutos a horas
CDC (Change Data Capture)	Captura y procesamiento de solo los cambios en los datos.	Replicación de cambios en una base de datos de ventas para análisis.	Segundos a minutos
Event Driven (Near Real Time)	Procesamiento casi en tiempo real desencadenado por eventos específicos.	Notificación de compra en una app de comercio electrónico.	Milisegundos a segundos
Asynchronous	Procesamiento donde el sistema no espera la finalización de una operación.	Envío de correo de verificación tras registro en una web.	Milisegundos a segundos
Real Time, Synchronous	Procesamiento en tiempo real donde el sistema espera la finalización de la operación.	Verificación de tarjeta de crédito en una compra online.	Milisegundos a menos de un segundo
Low Latency (Streaming)	Procesamiento de datos en flujo continuo con baja latencia.	Transmisión de datos de sensores en un coche autónomo.	Milisegundos

Conceptos esenciales



- Extraer, Transformar y Cargar
- Latencia

Conceptos esenciales

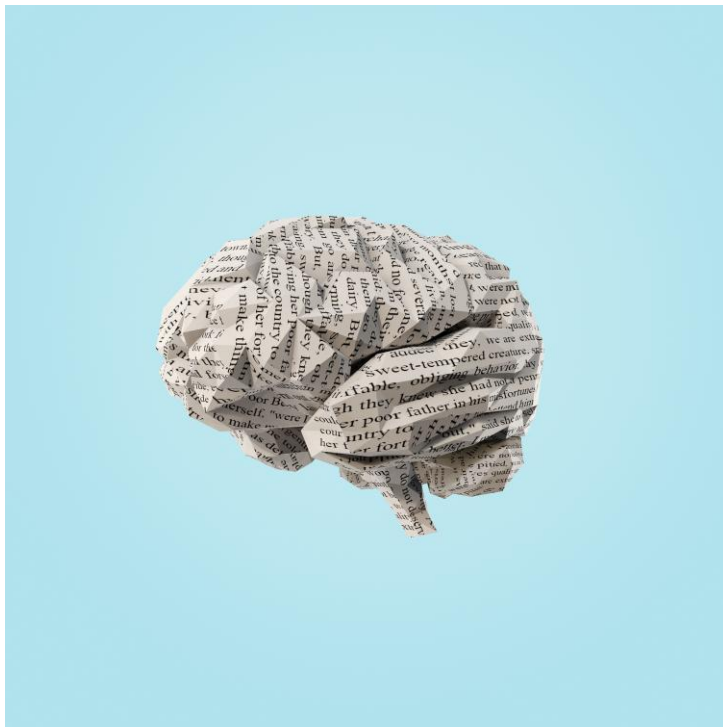


- Extraer, Transformar y Cargar
- Latencia
- Replicación

Replicación

- Para proporcionar un mejor tiempo de respuesta a los usuarios ubicados en todo el mundo, algunas aplicaciones mantienen **copias exactas de conjuntos de datos en múltiples ubicaciones físicas**.
- Las soluciones de replicación **minimizan el impacto del rendimiento** de los análisis y las consultas en el entorno operativo transaccional primario.
- Muchas de estas soluciones utilizan el **Change Data Capture** para vigilar los cambios en los diferentes sistemas y actuar de consecuencia “replicando” los solos los cambios efectivos.

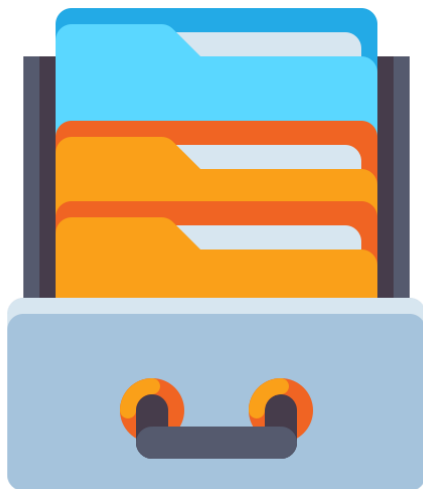
Conceptos esenciales



- Extraer, Transformar y Cargar
- Latencia
- Replicación

- Extraer, Transformar y Cargar
- Latencia
- Replicación
- Archivado

Archiving ⚠



Archiving



- Archivar os datos que se utilizan con poca frecuencia o que no se utilizan activamente.

Archiving



- Archivar os datos que se utilizan con poca frecuencia o que no se utilizan activamente.
- Las funciones ETL pueden utilizarse para transportar y posiblemente transformar los datos de archivo a las estructuras de datos del entorno de archivo.

Archiving



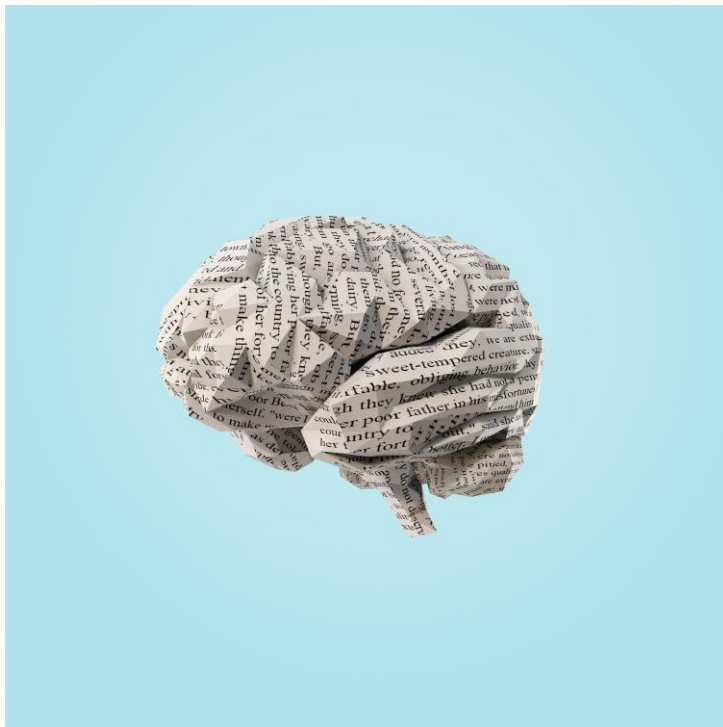
- Archivar os datos que se utilizan con poca frecuencia o que no se utilizan activamente.
- Las funciones ETL pueden utilizarse para transportar y posiblemente transformar los datos de archivo a las estructuras de datos del entorno de archivo.
- Mejora la eficiencia operativa

Archiving



- Archivar os datos que se utilizan con poca frecuencia o que no se utilizan activamente.
- Las funciones ETL pueden utilizarse para transportar y posiblemente transformar los datos de archivo a las estructuras de datos del entorno de archivo.
- Mejora la eficiencia operativa
- Supervisar antes cambios de tecnología de archivado.

Conceptos esenciales



- Extraer, Transformar y Cargar
- Latencia
- Replicación
- Archivado

- Extraer, Transformar y Cargar
- Latencia
- Replicación
- Archivado
- Modelo canónico

Modelo Canónico



- Un modelo de datos canónico es un **modelo común** utilizado por una organización o grupo de intercambio de datos que **estandariza el formato en el que se compartirán los datos**.
- En un patrón de diseño de interacción de datos de tipo "hub and spoke", todos los sistemas que quieren proporcionar o recibir datos interactúan sólo con un centro de información. (A veces llamado **DataHub**)
- Los datos se transforman desde o hacia un sistema emisor o receptor.

Conceptos esenciales



- Extraer, Transformar y Cargar
- Latencia
- Replicación
- Archivado
- Modelo canónico

Conceptos esenciales

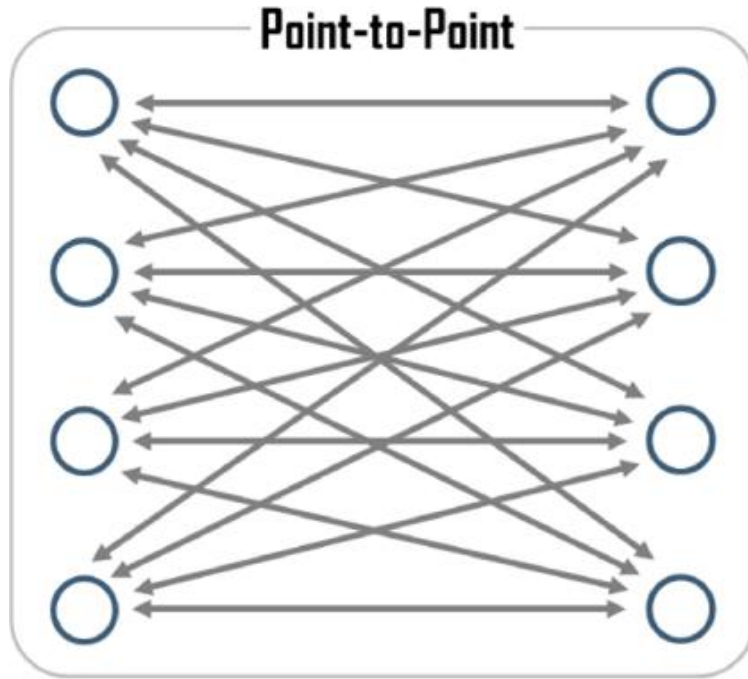


- Extraer, Transformar y Cargar
- Latencia
- Replicación
- Archivado
- Modelo canónico
- Modelos de interacción

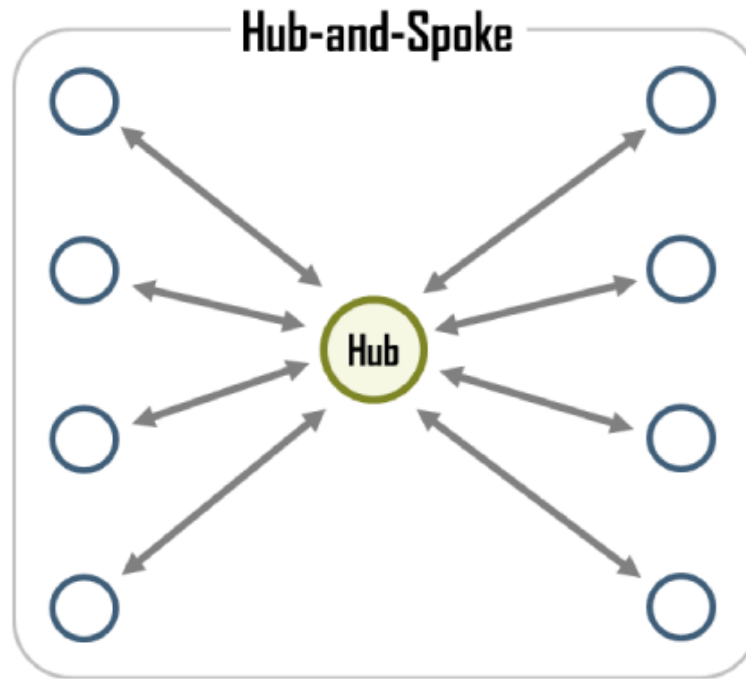
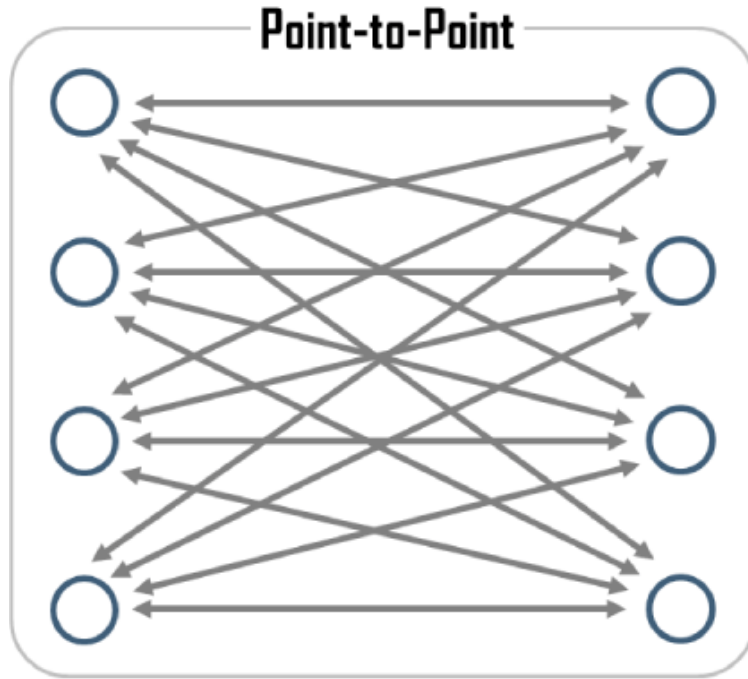
Modelos de interacción



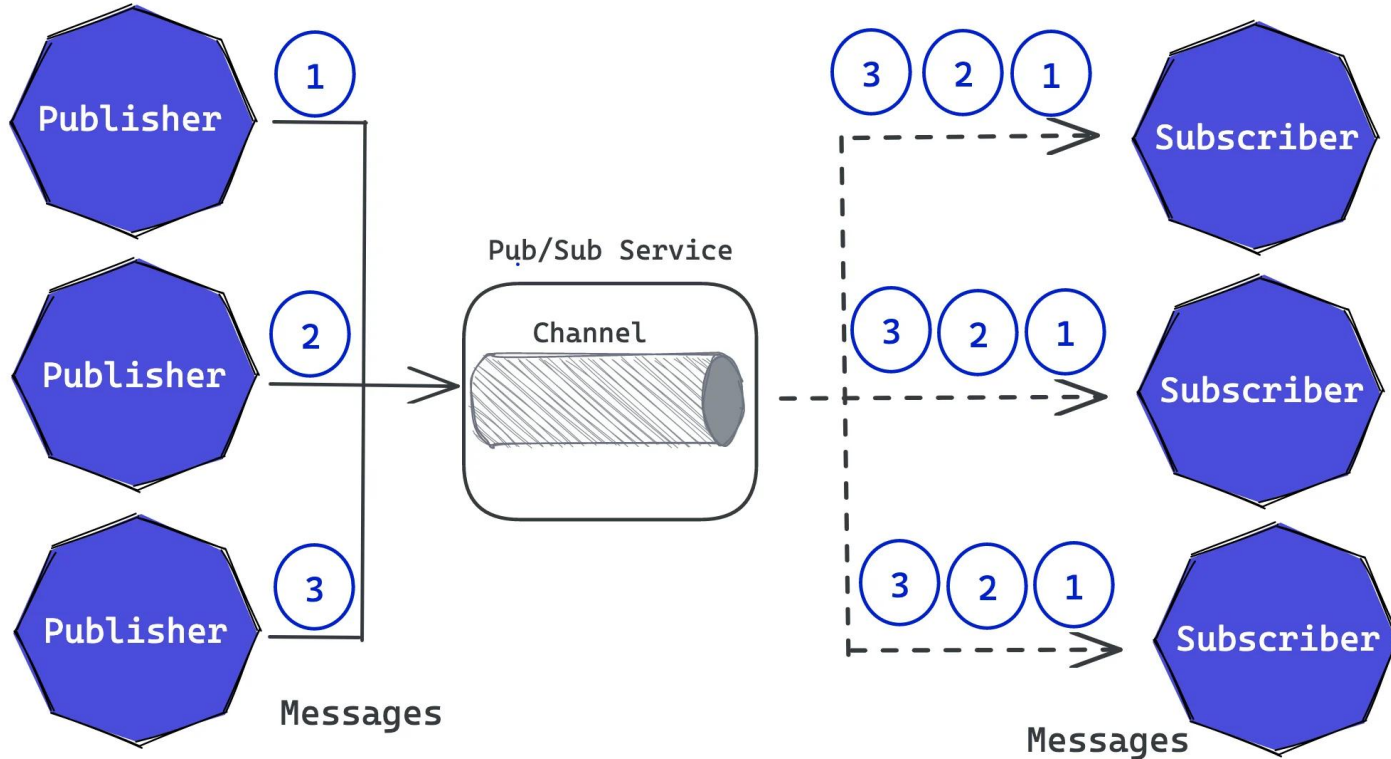
Modelos de interacción



Modelos de interacción



Publish-Subscribe



Conceptos esenciales



- Extraer, Transformar y Cargar
- Latencia
- Replicación
- Archivado
- Modelo canónico
- Modelos de interacción

Conceptos esenciales



- Extraer, Transformar y Cargar
- Latencia
- Replicación
- Archivado
- Modelo canónico
- Modelos de interacción
- Arquitecturas de integración e interoperabilidad

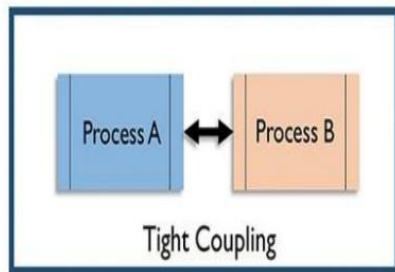
Arquitectura de integración e interoperabilidad.

Conceptos

- *Application Coupling*
- *Orchestration and Process Control*
- *Enterprise Application Integration (EAI)*
- *Enterprise Service Bus (ESB)*
- *Service-Oriented Architecture (SOA)*
- *Complex Event Processing (CEP)*
- *Data Federation & Virtualization*
- *Data-as-Service (DaaS)*
- *Cloud Based Integration*

Arquitectura de integración e interoperabilidad.

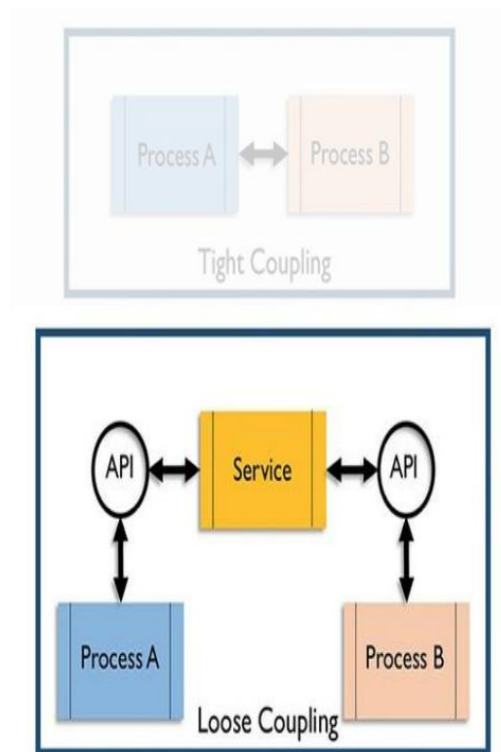
Acoplamiento de aplicaciones: Tightly Coupled



- El acoplamiento describe el grado en que dos sistemas **están entrelazados**.
- Dos sistemas fuertemente acoplados suelen tener una **interfaz sincrónica**, en la que un sistema espera la respuesta del otro.
- El **acoplamiento estrecho** representa una operación más arriesgada: si uno de los sistemas no está disponible, los dos no lo están
- El plan de continuidad de la actividad tiene que ser el mismo para ambos sistemas

Arquitectura de integración e interoperabilidad.

Acoplamiento de aplicaciones: Loosed Coupled



- Siempre que sea posible, el acoplamiento flexible es un diseño de interfaz preferido, en el que los datos se transmiten entre sistemas sin esperar una respuesta y un sistema puede no estar disponible sin que el otro no esté disponible.
- El acoplamiento flexible se puede implementar mediante varias técnicas con servicios, API o colas de mensajes.
- La arquitectura orientada a servicios mediante un bus de servicios empresariales es un ejemplo de un patrón de diseño de interacción de datos de acoplamiento flexible.

Arquitectura de integración e interoperabilidad.

Orquestación y control de procesos



- **Orquestación** es el término utilizado para describir cómo se organizan y ejecutan múltiples procesos en un sistema.

Arquitectura de integración e interoperabilidad.

Orquestación y control de procesos



- **Orquestación** es el término utilizado para describir cómo se organizan y ejecutan múltiples procesos en un sistema.
- **Los controles de procesos** son los componentes que garantizan que el envío, la entrega, la extracción y la carga de datos sean precisos y completos.

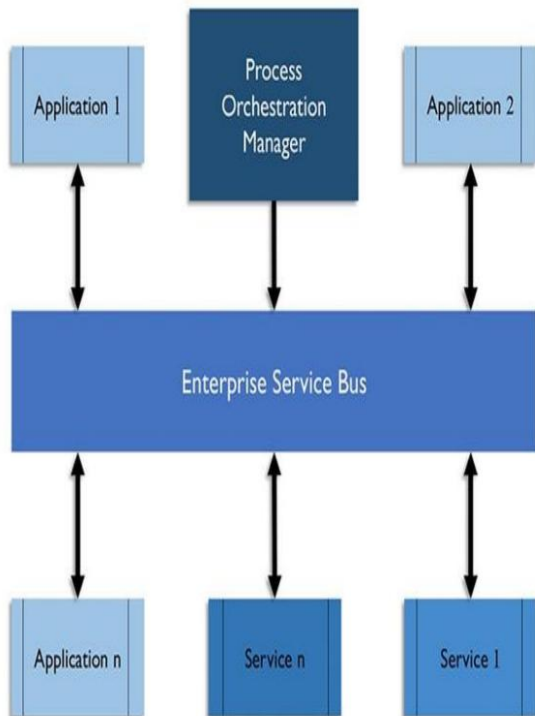
Arquitectura de integración e interoperabilidad

Integración de aplicaciones empresariales (EAI)

- En un modelo de integración de aplicaciones empresariales (EAI), los módulos de software interactúan entre sí **únicamente a través de llamadas de API**.
- **Los almacenes de datos son actualizados únicamente por sus propios módulos de software** y otro software no puede llegar a los datos de una aplicación, sino que sólo accede a través de las API definidas.
- La EAI se basa en conceptos orientados a los objetos, que hacen hincapié en la reutilización y en la posibilidad de sustituir cualquier módulo sin que ello afecte a ningún otro.
- Tiene un enfoque absolutamente holístico

Arquitectura de integración e interoperabilidad.

Bus de servicios empresariales (ESB)



- Un *Enterprise Service Bus* es un sistema que actúa como intermediario entre sistemas, pasando mensajes entre ellos.
- Las aplicaciones pueden enviar y recibir mensajes o archivos utilizando el ESB, y están encapsuladas de otros procesos existentes en el ESB.
- ESB actúa como servicio entre las aplicaciones

Arquitectura de integración e interoperabilidad.

Service-Oriented Architecture (SOA)

Arquitectura de integración e interoperabilidad.

Service-Oriented Architecture (SOA)

- Las aplicaciones no tienen que interactuar directamente con otras aplicaciones ni conocer su funcionamiento interno.

Arquitectura de integración e interoperabilidad.

Service-Oriented Architecture (SOA)

- Las aplicaciones no tienen que interactuar directamente con otras aplicaciones ni conocer su funcionamiento interno.
- SOA permite la **independencia de las aplicaciones** y la posibilidad de que una organización sustituya los sistemas sin necesidad de realizar cambios significativos en los sistemas

Arquitectura de integración e interoperabilidad.

Service-Oriented Architecture (SOA)

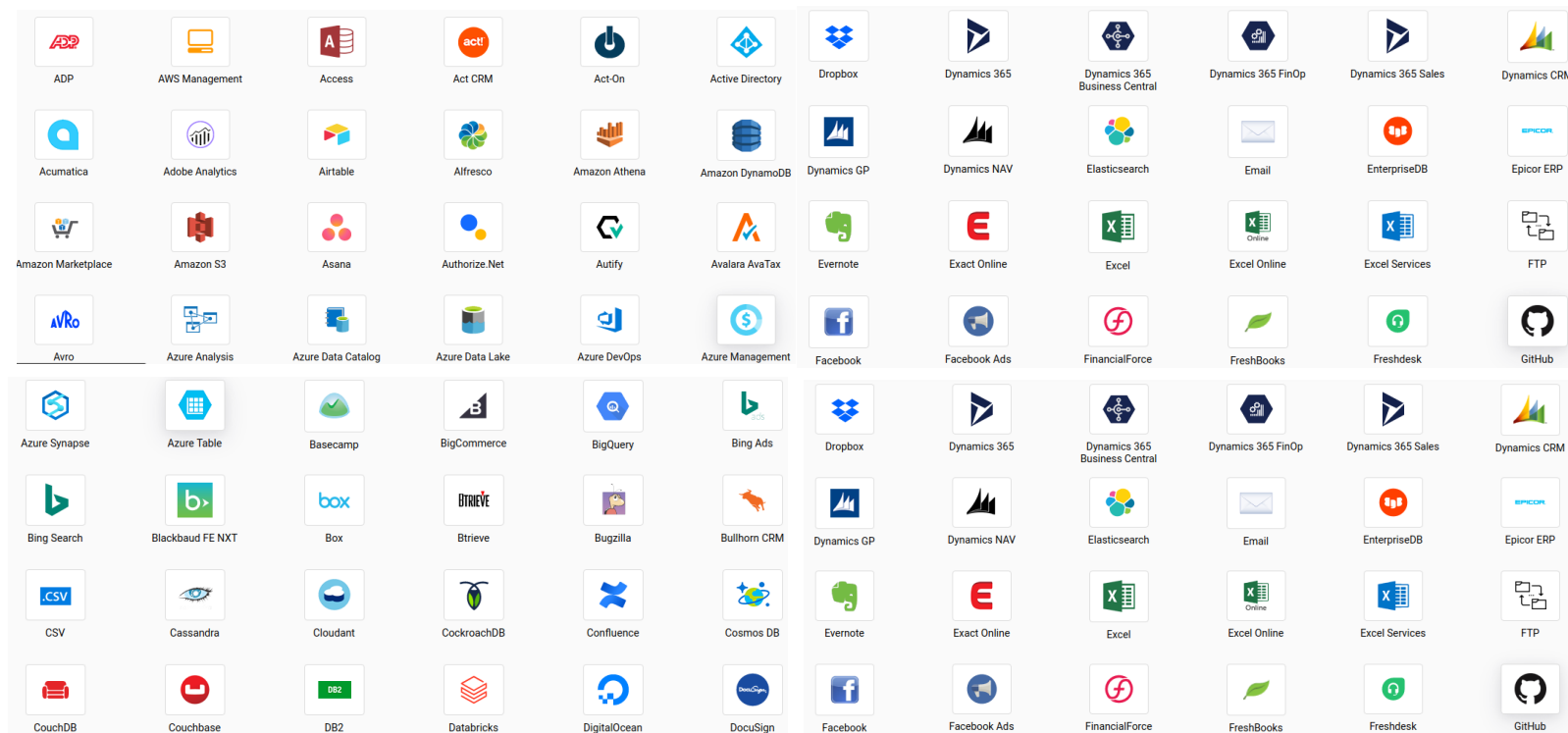
- Las aplicaciones no tienen que interactuar directamente con otras aplicaciones ni conocer su funcionamiento interno.
- SOA permite la **independencia de las aplicaciones** y la posibilidad de que una organización sustituya los sistemas sin necesidad de realizar cambios significativos en los sistemas
- Una SOA se puede implementar con varias tecnologías, incluidos servicios web, mensajería, API RESTful, etc. **Por lo general, los servicios se implementan como API** (interfaces de programación de aplicaciones) que están disponibles para ser llamadas por los sistemas de aplicaciones (o consumidores humanos)

Arquitectura de integración e interoperabilidad.

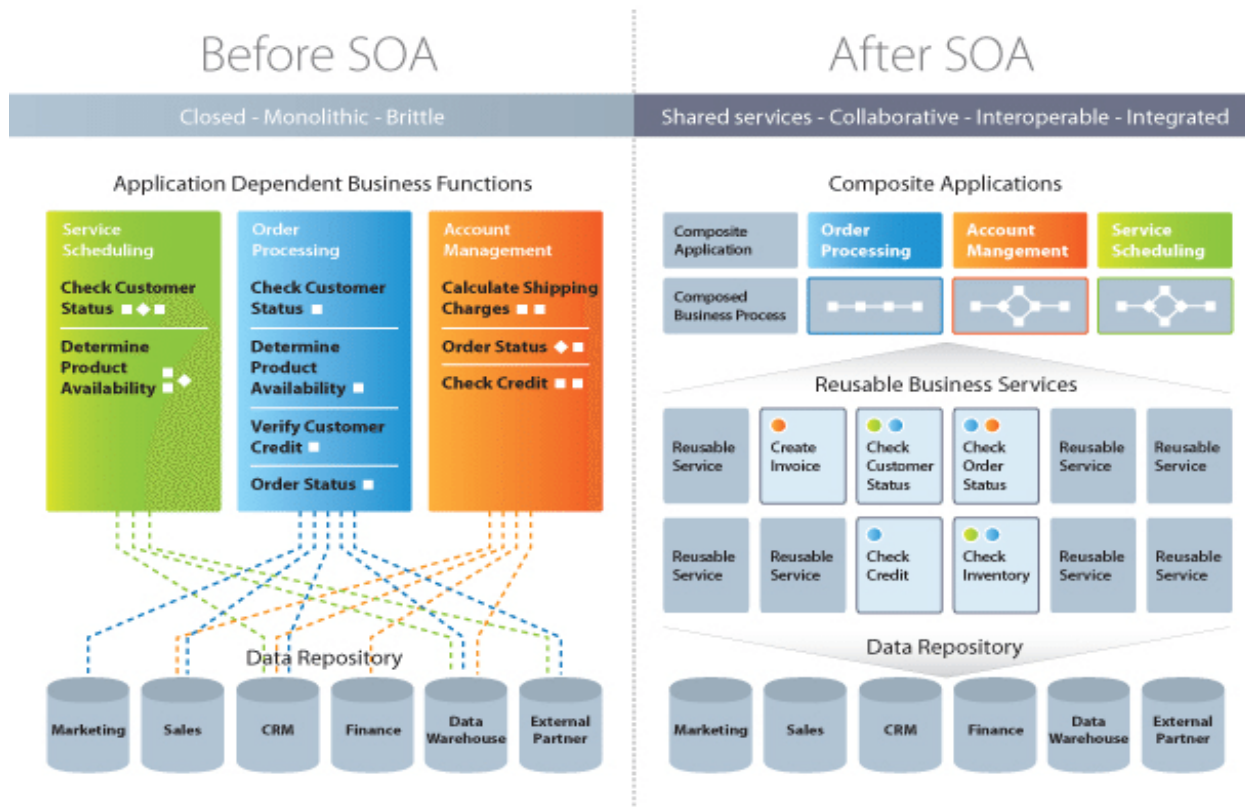
Service-Oriented Architecture (SOA)

- Las aplicaciones no tienen que interactuar directamente con otras aplicaciones ni conocer su funcionamiento interno.
- SOA permite la **independencia de las aplicaciones** y la posibilidad de que una organización sustituya los sistemas sin necesidad de realizar cambios significativos en los sistemas
- Una SOA se puede implementar con varias tecnologías, incluidos servicios web, mensajería, API RESTful, etc. **Por lo general, los servicios se implementan como API** (interfaces de programación de aplicaciones) que están disponibles para ser llamadas por los sistemas de aplicaciones (o consumidores humanos)

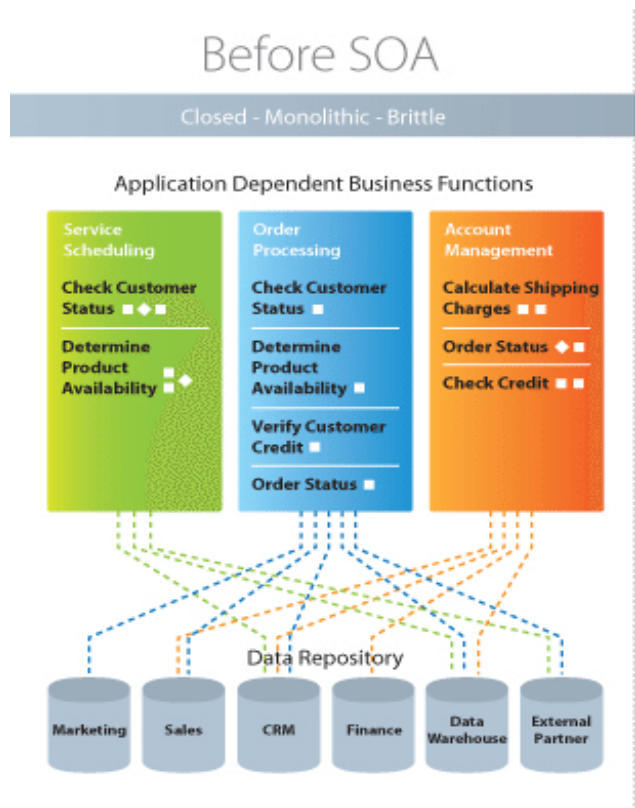
Un mundo de API, un mundo de posibilidades



Un mundo de API, un mundo de posibilidades

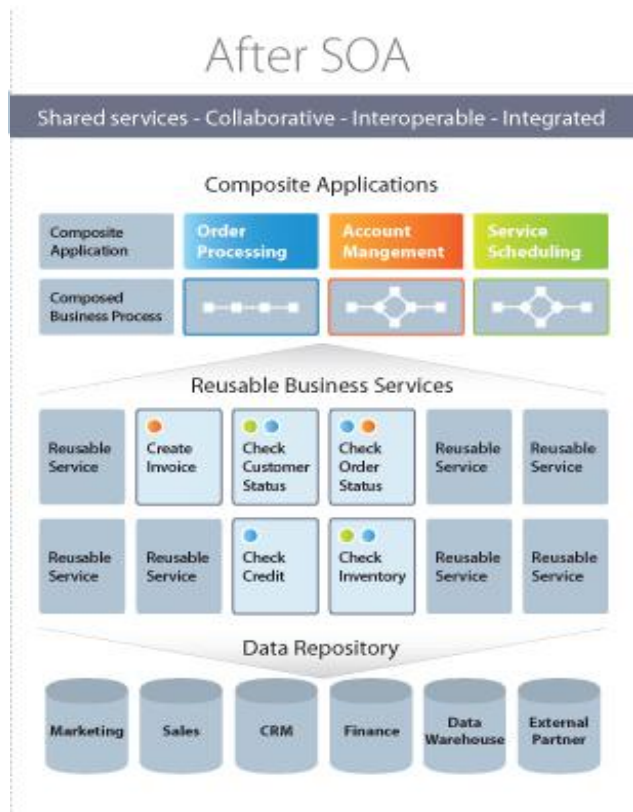


Un mundo de API, un mundo de posibilidades



- **Sistemas Monolíticos**
 - Estructura Rígida
 - Dependencia del Hardware y del S.O
- **Integración Costosa y Lenta**
 - Integración Punto a Punto
 - Mantenimiento Difícil
- **Falta de Reusabilidad**
 - Duplicación de Funcionalidad
- **Barreras para la Agilidad Empresarial**
 - Lanzamientos Lentos

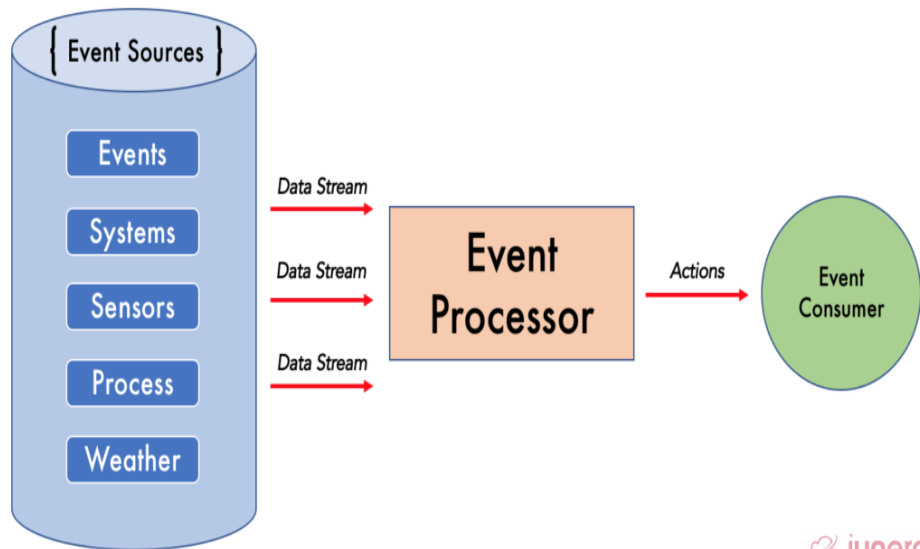
Un mundo de API, un mundo de posibilidades



- **Descomposición en Servicios:**
 - Modularidad
 - Independencia del Hardware
- **Integración Eficiente:**
 - Interoperabilidad Mejorada
 - Gestión Centralizada
- **Reusabilidad de Componentes:**
 - Servicios Reutilizables
- **Agilidad Empresarial:**
 - Rápida Respuesta a Cambios

Arquitectura de integración e interoperabilidad.

Complex Event Processing (CEP)



- El procesamiento de eventos es un método para rastrear y analizar (procesar) flujos de información (datos) sobre cosas que suceden (eventos),
- El procesamiento de eventos complejos (CEP) combina datos de múltiples fuentes para identificar eventos significativos (como oportunidades o amenazas) para predecir el comportamiento o la actividad y desencadenar automáticamente una respuesta en tiempo real.
- Se establecen reglas para guiar el procesamiento y el enrutamiento de los eventos.

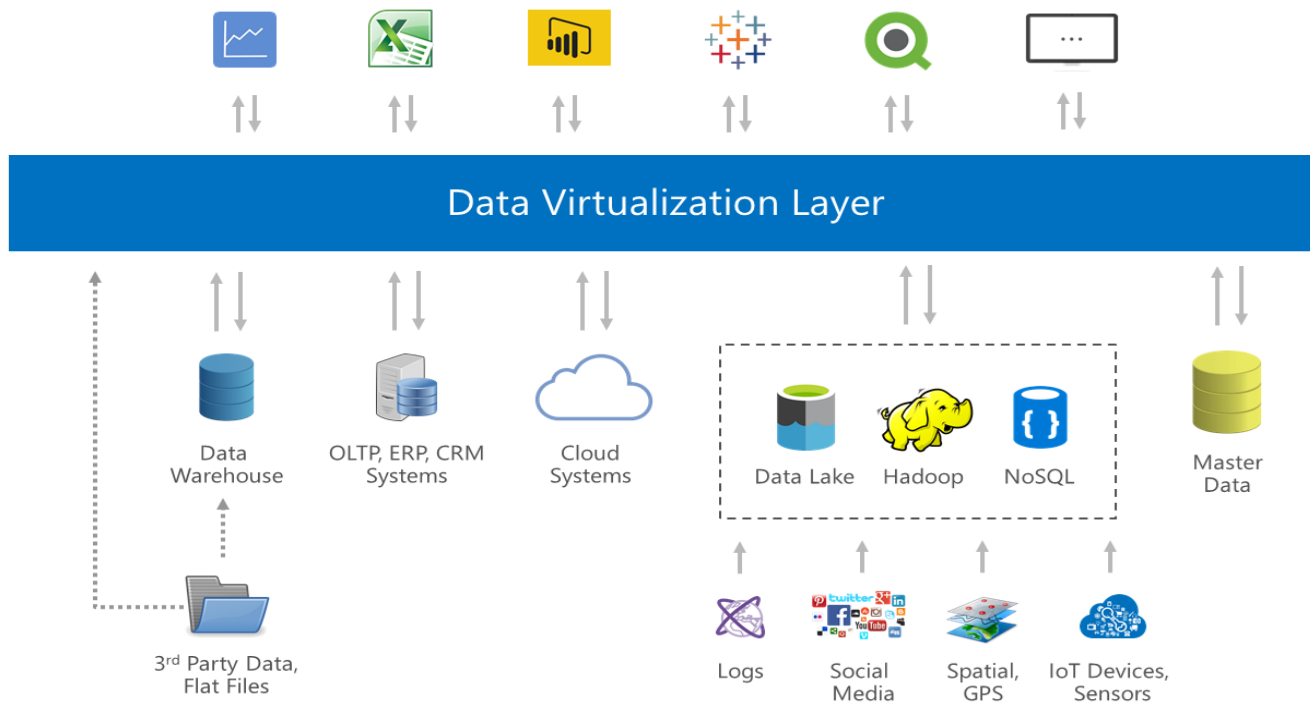
Arquitectura de integración e interoperabilidad.

Data Virtualization & Federation

.

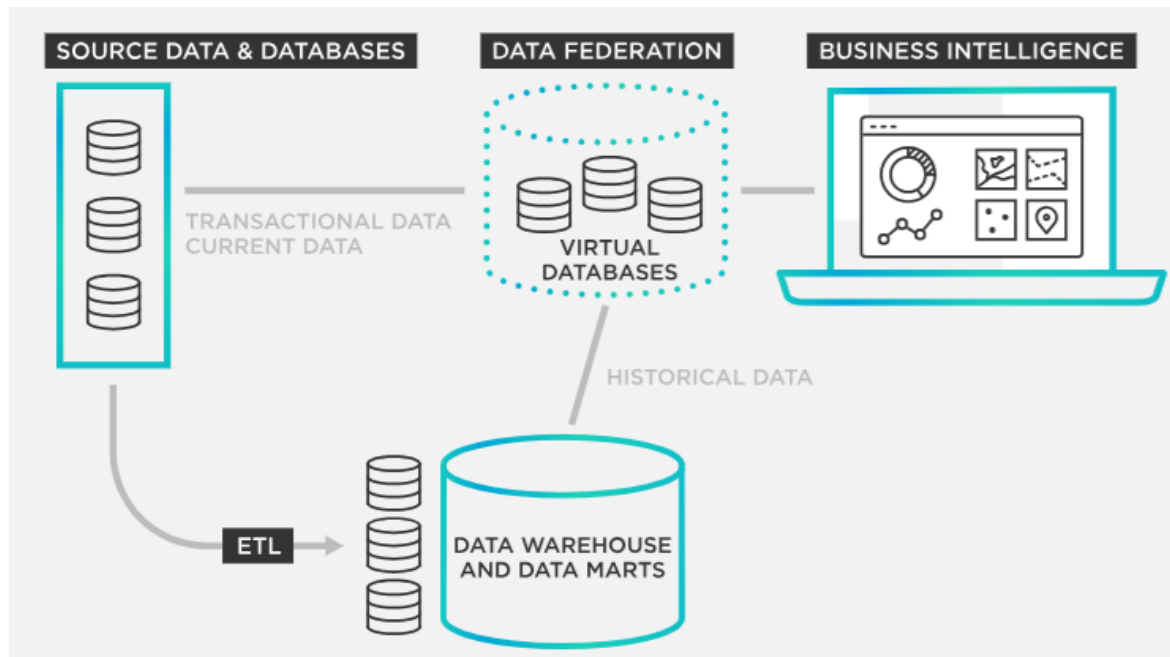
Arquitectura de integración e interoperabilidad.

Data Virtualization



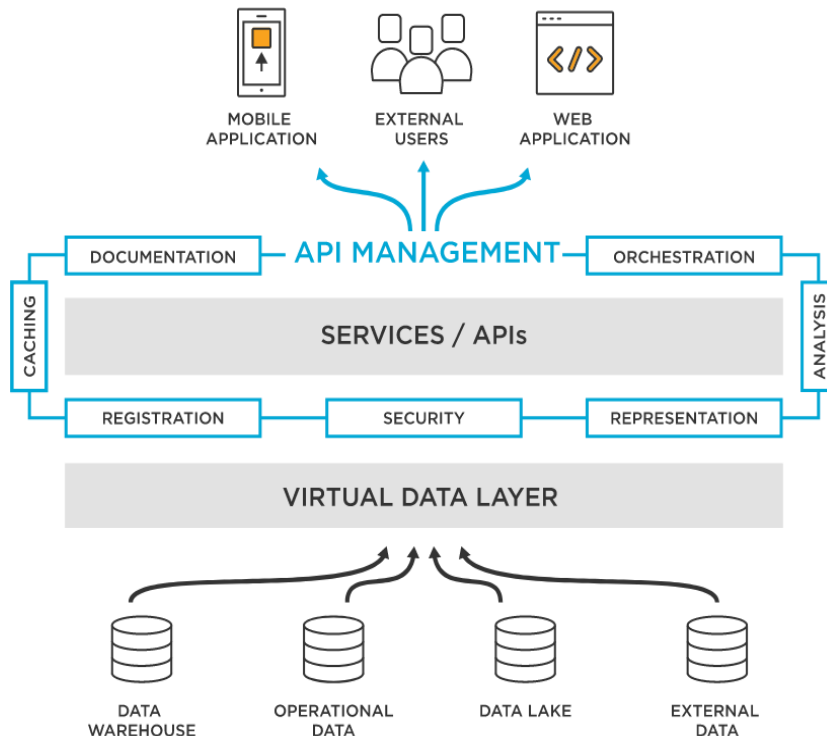
Arquitectura de integración e interoperabilidad.

Data Federation



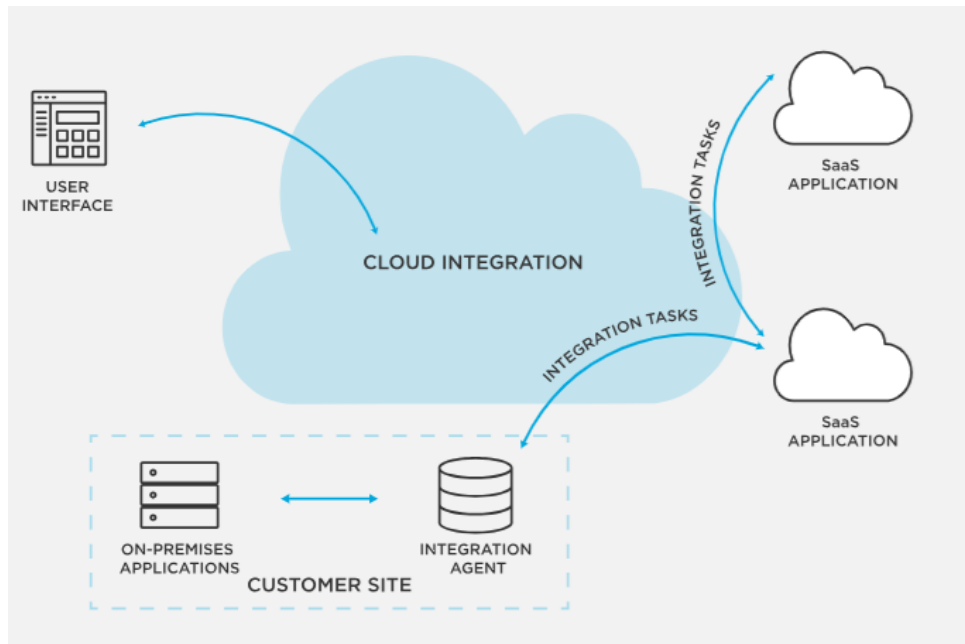
Arquitectura de integración e interoperabilidad.

Data as a Service



Arquitectura de integración e interoperabilidad.

Cloud Based Integration



La integración basada en la nube (también conocida como plataforma de integración como servicio o **IPaaS**) es una forma de integración de sistemas suministrada como servicio en la nube que aborda casos de uso de integración de datos, procesos, arquitectura orientada a servicios (SOA) y aplicaciones.

Herramientas de DII



Data Transformation Engine/ETL Tool

Herramientas de DII



Data Transformation Engine/ETL Tool



Data Virtualization Server

Herramientas de DII



Data Transformation Engine/ETL Tool



Data Virtualization Server



Enterprise Service Bus

Herramientas de DII



Data Transformation Engine/ETL Tool



Data Virtualization Server



Enterprise Service Bus



Business Rules Engine

Herramientas de DII



Data Transformation Engine/ETL Tool



Data Virtualization Server



Enterprise Service Bus



Business Rules Engine



Data and Process Modeling Tools

Herramientas de DII



Data Transformation Engine/ETL Tool



Data Virtualization Server



Enterprise Service Bus



Business Rules Engine

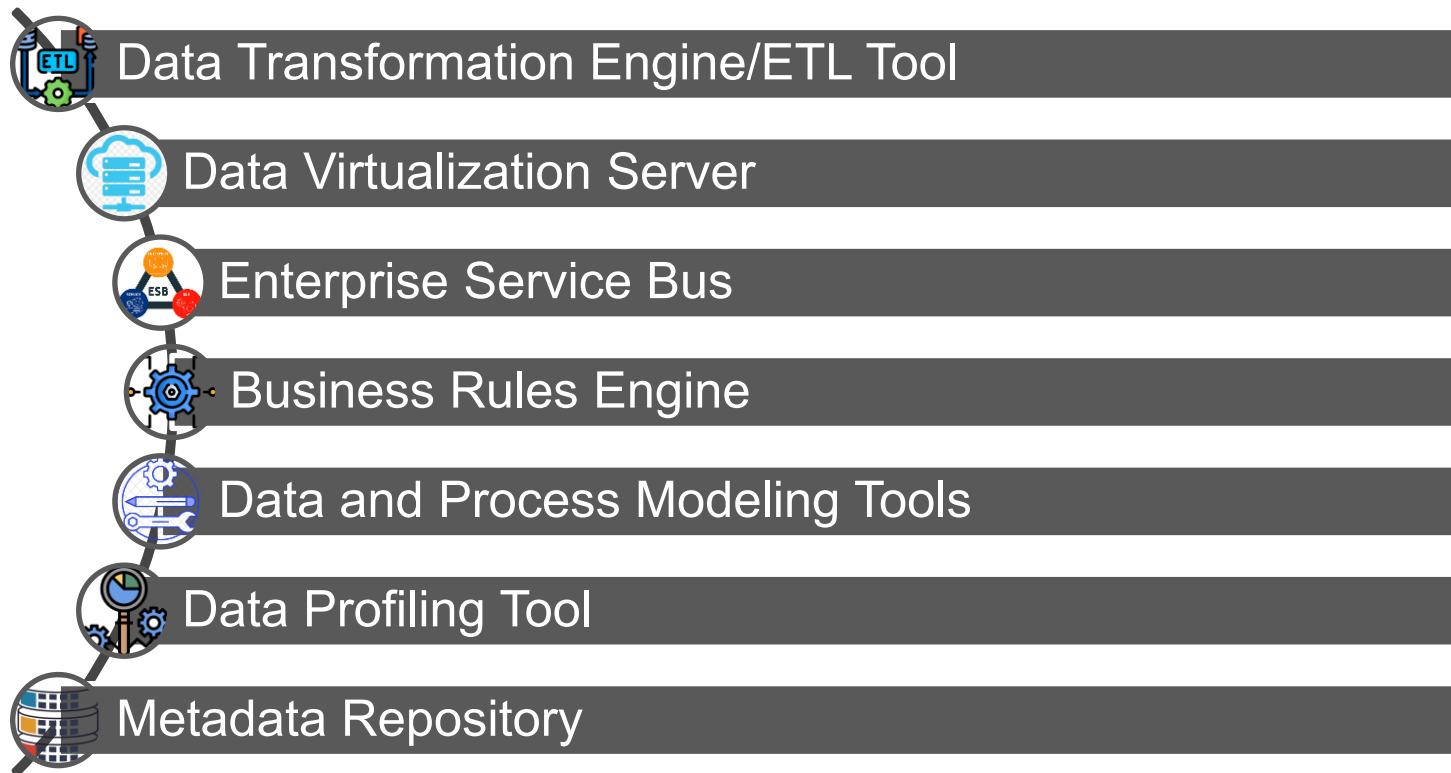


Data and Process Modeling Tools



Data Profiling Tool

Herramientas de DII



DII Governance

- Las decisiones sobre el diseño de mensajes de datos, modelos de datos y reglas de transformación de datos tienen un impacto directo en la capacidad de una organización para usar sus datos. **Estas decisiones deben estar impulsadas por el negocio.**

DII Governance

- Las decisiones sobre el diseño de mensajes de datos, modelos de datos y reglas de transformación de datos tienen un impacto directo en la capacidad de una organización para usar sus datos. **Estas decisiones deben estar impulsadas por el negocio.**
- Negocio es responsable de:
 - Definir las reglas sobre cómo se deben modelar y transformar los datos.
 - Aprobar los cambios en cualquiera de estas reglas de negocio.

DII Governance

Data Sharing Agreements



DII Governance

Data Sharing Agreements



- Establecer MOU (*memorandum of understanding*) aprobado por los Data Stewards.

DII Governance

Data Sharing Agreements



- Establecer MOU (*memorandum of understanding*) aprobado por los Data Stewards.
- Los acuerdos de intercambio de datos deben especificar el uso y el acceso anticipados a los datos, las restricciones de uso, así como los niveles de servicio esperados.

DII Governance

Data Sharing Agreements



- Establecer MOU (*memorandum of understanding*) aprobado por los Data Stewards.
- Los acuerdos de intercambio de datos deben especificar el uso y el acceso anticipados a los datos, las restricciones de uso, así como los niveles de servicio esperados.
- Especial relevancia en sectores regulados o cuando se gestionan información de carácter personal.

DII Governance

DII and Data Lineage



DII Governance

DII and Data Lineage



- La gobernanza es necesaria para garantizar que se documente el conocimiento de los orígenes y el movimiento de los datos.

DII Governance

DII and Data Lineage



- La gobernanza es necesaria para garantizar que se documente el conocimiento de los orígenes y el movimiento de los datos.
- Existen estándares de cumplimiento emergentes (por ejemplo, la regulación de Solvencia II en Europa) que requieren que las organizaciones puedan describir dónde se originaron sus datos y cómo se han modificado a medida que se han movido a través de varios sistemas.

DII Governance

DII and Data Lineage



- La gobernanza es necesaria para garantizar que se documente el conocimiento de los orígenes y el movimiento de los datos.
- Existen estándares de cumplimiento emergentes (por ejemplo, la regulación de Solvencia II en Europa) que requieren que las organizaciones puedan describir dónde se originaron sus datos y cómo se han modificado a medida que se han movido a través de varios sistemas.
- El linaje de datos es fundamental como parte del análisis de impacto necesario al realizar cambios en las estructuras de datos, los flujos de datos o el procesamiento de datos.

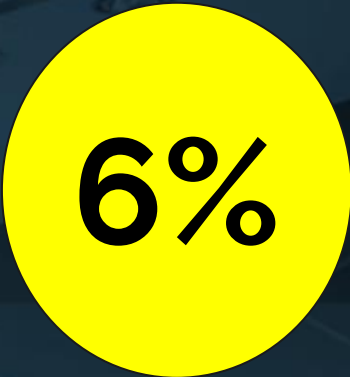
DII Governance

Métricas de Integración de datos



- **Disponibilidad de Datos.**
 - Disponibilidad de los datos requeridos.
- **Volúmenes de Datos y Velocidad.**
 - Volúmenes de datos transportados y transformados.
 - Volúmenes de datos analizados.
 - Velocidad de transmisión.
 - Latencia entre actualización y disponibilidad de datos.
 - Latencia entre el evento y la acción desencadenada.
 - Tiempo de disponibilidad de nuevas fuentes de datos.
- **Costos y Complejidad de la Solución.**
 - Costo de desarrollo y de gestión de soluciones
 - Facilidad para adquirir nuevos datos
 - Complejidad de soluciones y operaciones
 - Número de sistemas que utilizan soluciones de integración de datos

Questions & Answers



6%

Examples of interaction models include:

- A. Hub-and-spoke
- B. Publish - subscribe
- C. Point-to-point
- D. Wheel-and-spike

The goals of Data Integration and Interoperability include:

- A. Provide data securely, with regulatory compliance, in the format and time frame needed.
- B. Lower cost and complexity of managing solutions by developing shared models and interfaces.
- C. Managing the availability of data throughout the data lifecycle
- D. Provide the starting point for customizations, integration or even replacement of an application
- E. Identify meaningful events and automatically trigger alerts and actions.
- F. Support business intelligence, analytics, master data management and operational efficiency efforts.

Data Integration and Interoperability is dependent on these other areas of data management:

- A. Metadata
- B. Data architecture
- C. Data governance
- D. Data security
- E. Data modelling and design
- F. Data storage and operations
- G. Data Quality

Latency can be:

- A. Batch
- B. Event-driven
- C. Distributed
- D. Real-time synchronous
- E. None of the above

ETL is:

- A. Event Time Latency
- B. Extract Transform Latency
- C. Extract Transform Load
- D. Event Transform Latency

ETL and ELT are the Same

- A. True
- B. False

¡Muchas gracias!



GOLD SPONSORS



SILVER SPONSORS



Entidades académicas

