

Gobierno del Dato y Toma de Decisiones

Tema 4. Metodologías y tendencias

Índice

Esquema

Ideas clave

- 4.1. Introducción y objetivos
- 4.2. Metodología Kimball
- 4.3. Metodologías PMI
- 4.4. Metodología Inmon
- 4.5. Data-driven decision modelling
- 4.6. Metodología DevOps
- 4.7. Nuevos roles
- 4.8. Tendencias
- 4.9. Referencias bibliográficas

A fondo

Metodología de Ralph Kimball para la implementación de DW/BI

El concepto de DevOps

Business intelligence governance best practices for 2021 and beyond

Test

METODOLOGÍAS Y TENDENCIAS	
Metodología Kimball	Principios básicos del ciclo. Rutas de la metodología.
Metodología PMI	Ejecución de procesos. Áreas de conocimiento.
Metodología Inmon	Características del DW. Se debe tener presente.
<i>Data-driven</i>	Mejora de proceso de toma de decisiones.
Tendencias	Análisis de las últimas tendencias según los expertos en BI y gobierno del dato.

4.1. Introducción y objetivos

El tema de las metodologías es quizá el más importante, ya que, si no se tiene en cuenta los pasos que se deben seguir para implementar cualquier estrategia BI o BA, el proyecto tiende al fracaso. En este apartado presentamos las metodologías que han sido el cimiento de este tipo de estrategias y algunas que se han adaptado con el tiempo y se consideran más modernas.

Los objetivos de este tema son:

- ▶ Conocer las metodologías clásicas del BI, la de Inmon y la de Kimball, reconocer sus ventajas y desventajas.
- ▶ Utilizar metodologías aplicadas a otras temáticas y aplicarlas a la estrategia BI
- ▶ Profundizar en lo que es una empresa *data-driven*.
- ▶ Analizar las tendencias en el campo del BI y del gobierno del dato.
- ▶ Descubrir los nuevos roles que se están creando para poder afrontar los nuevos retos.

4.2. Metodología Kimball

Existen varias metodologías de diseño y construcción del *data warehouse* (DW). Cada productor de *software* de *business intelligence* intenta aplicar una metodología con sus productos. Las metodologías de Kimball y de Inmon son las más utilizadas actualmente. Para entender la diferencia entre estas dos metodologías, volvamos a los conceptos de ***data warehouse*** y de ***data mart***. El segundo es un repositorio de información, similar a un DW, pero más pequeño y orientado a un área o departamento específico de la organización (por ejemplo: Compras, Ventas o RR. HH.), a diferencia del DW, que cubre toda la organización y su alcance (Kimball et al., 1998).

La metodología de Kimball se centra principalmente en el diseño de la base de datos que guardará la información para la toma de decisiones. El diseño se basa en la creación de tablas de hechos, que contienen la información numérica de los indicadores por analizar, es decir, la parte cuantitativa de la información.

La aplicación de cualquier metodología para el desarrollo del DW en las empresas siempre va a depender de las necesidades de las organizaciones y el compromiso institucional de quienes conforman dichas organizaciones.

La metodología se basa en lo que Kimball denomina ciclo de vida dimensional del negocio —*business dimensional lifecycle*— (Kimball et al., 1998; Mundy et al., 2006).

El ciclo de vida del proyecto de DW está fundamentado en cuatro principios básicos:

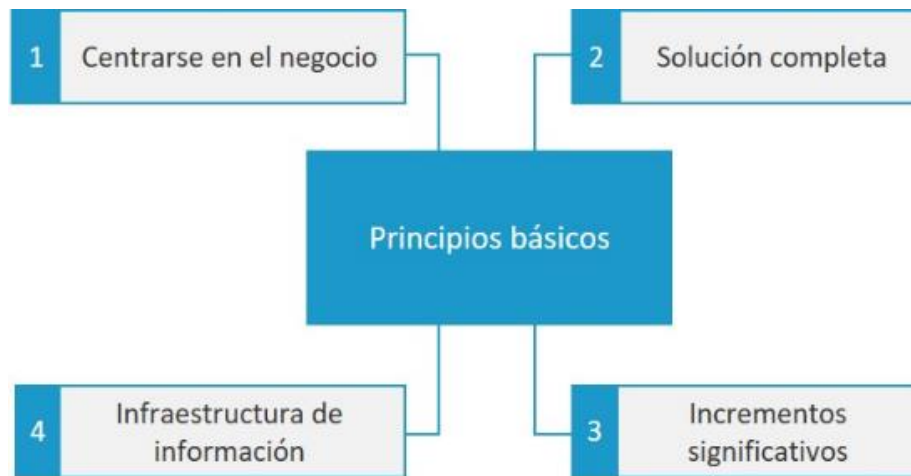


Figura 1. Principios básicos para el desarrollo del DW, según la metodología Kimball. Fuente: Kimball et al., 1998.

- 1. Centrarse en el negocio:** es necesario enfocarse hacia la tipificación de los requerimientos del negocio y el valor que le aportará, emplear estos esfuerzos en desarrollar relaciones sólidas con el negocio y con los encargados de implementar las herramientas.
- 2. Realizar una infraestructura de información:** se trata de diseñar una base de información única, integrada, fácil de usar y de alto rendimiento, en la cual deben verse reflejados todos los requerimientos de negocio identificados en la organización.
- 3. Cumplir las entregas en incrementos significativos:** define los tiempos de creación del almacén de datos en incrementos progresivos entregables. Los plazos están determinados por los desarrolladores del proyecto y estos pueden ser cuatro o seis meses cada entrega (ten en cuenta que estos son tan solo ejemplos, los tiempos de entrega se deben ajustar a las necesidades de la empresa).
- 4. Ofrecer la solución completa:** facilitar todos los elementos necesarios para entregar valor a todos los usuarios del negocio. Para empezar, se debe contar con un almacén de datos sólido, bien diseñado, con calidad probada y accesible. También

se deberá entregar herramientas de consulta *ad hoc*, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación (Rivadera, 2010).

La construcción de una solución de *data warehouse/business intelligence* no es una tarea fácil; sin embargo, Kimball propone una metodología que ayuda a facilitar esa complejidad. Las tareas de esta metodología (ciclo de vida) se muestran en la Figura 2.

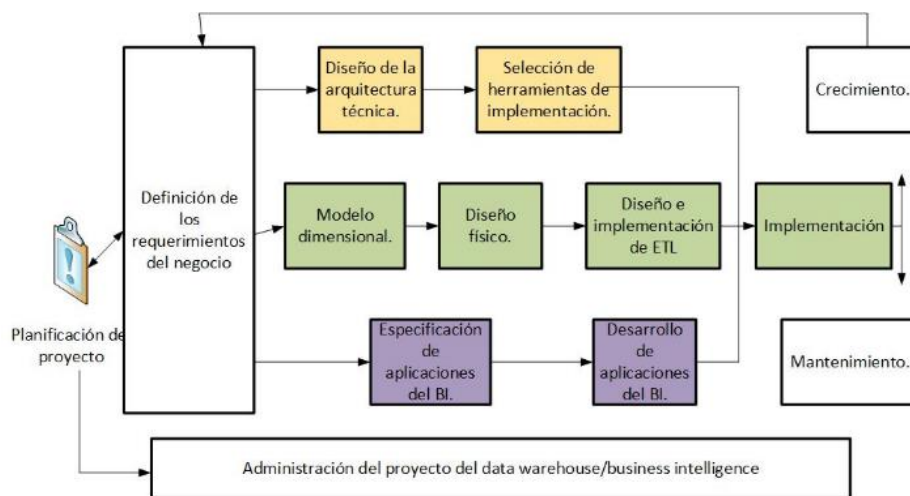


Figura 2. Tareas de la metodología de Kimball, denominada business dimensional lifecycle. Fuentes: Kimball et al., 1998; Mundy et al., 2006.

En la Figura 2, se observa que hay que destacar el papel central de la tarea de definición de requerimientos. Los requerimientos del negocio son la base inicial de las tareas siguientes. En segundo lugar, se observan tres rutas que se enfocan en tres áreas diferentes:

- ▶ **Tecnología (ruta superior):** envuelve tareas relacionadas con herramientas *software* específicas para cada empresa, por ejemplo: *data integration* —Kettle — Pentaho Open Source.
- ▶ **Datos (ruta interior):** se diseñará e implementará el modelo dimensional y desarrollo del subsistema de ETL (extracción, transformación y carga) para cargar el

data warehouse.

- ▶ **Aplicaciones del *business intelligence* (ruta inferior):** en esta ruta se centran las tareas para el diseño, desarrollo e implementación de las aplicaciones de negocio para los usuarios finales.

Estas rutas se van mezclando a medida que se va realizando el proyecto hasta llegar a la instalación del sistema. En la base de la Figura 2, se observa la actividad general de administración del proyecto. A continuación, se amplía la información de cada una de las tareas.

Planificación del proyecto

En esta parte se establece el propósito del proyecto de *data warehouse/business intelligence*, los objetivos específicos, el alcance del mismo, los riesgos y un acercamiento a las necesidades de información. Para Kimball et al. (1998), el proyecto hace referencia a una iteración simple del KLC (*Kimball Life Cycle*), desde el principio hasta el despliegue. Esta tarea incluye estas fases comunes para un plan de proyecto:

- ▶ Definir lo más certero posible el alcance (requerimientos del negocio).
- ▶ Identificar las tareas.
- ▶ Programar las tareas.
- ▶ Planear el uso de los recursos.
- ▶ Establecer la carga de trabajo a los diferentes recursos.
- ▶ Preparar una prueba piloto.
- ▶ Preparar y elaborar el documento final (informes).

En el apartado de la planeación de los recursos humanos, Kimball propone la

conformación de un equipo de trabajo al que pertenece tanto personal del área de informática como personal directivo (expertos en el negocio). En la Tabla 1 se muestran algunos ejemplos.

RECURSOS HUMANOS	
Negocios	Tecnologías de la información
Cliente final	Arquitecto técnico
Gerente	Modelador de datos
Líder	Coordinador de <i>data marts</i> y metadatos
Usuarios	Administrador de bases de datos
Expertos de áreas	Soporte del DW

Tabla 1. Ejemplo de recursos humanos según metodología Kimball. Fuente: Kimball et al., 1998.

Análisis de requerimientos

Para poder conseguir los requerimientos del negocio, se debe planear cómo obtendremos dichos requerimientos, y para ello existen dos técnicas principales:

- ▶ Las entrevistas.
- ▶ Las sesiones facilitadoras.

Las entrevistas deben realizarse a los expertos en el negocio (**directivos**) y a los **técnicos**. Para ello es preciso tener una preparación previa con ayuda de un experto. De este proceso es imperativo aprender todo sobre el negocio, la competencia, la industria y los clientes (internos y externos), así como conocer las pautas y la terminología. A partir de este análisis se puede construir una herramienta, denominada **matriz de procesos/dimensiones**, *bus matrix* en inglés (Rivadera,

2010).

La dimensión es una **forma** o **criterio** por medio del cual se pueden resumir, cruzar o cortar datos numéricos para ser analizados. La matriz generada contiene en las filas los procesos de negocios detallados y, en las columnas, las dimensiones reconocidas. Un ejemplo de esta matriz de ventas se puede observar en la Tabla 2:

Dimensiones							
Hechos		Fechas	Artículo	Almacenes	Empleados	Clientes	Proveedores
	Ventas	X	X	X	X	X	
	Inventarios	X	X	X			
	Compras	X	X		X		X
	Facturación	X	X		X		

Tabla 2. Ejemplo de matriz hecho/dimensiones (*bus matrix*). Fuente: Zorrilla, 2011.

Modelo dimensional

La creación de un modelo dimensional es un proceso eficiente y muy iterativo. En la Figura 3 se aprecia un esquema general.

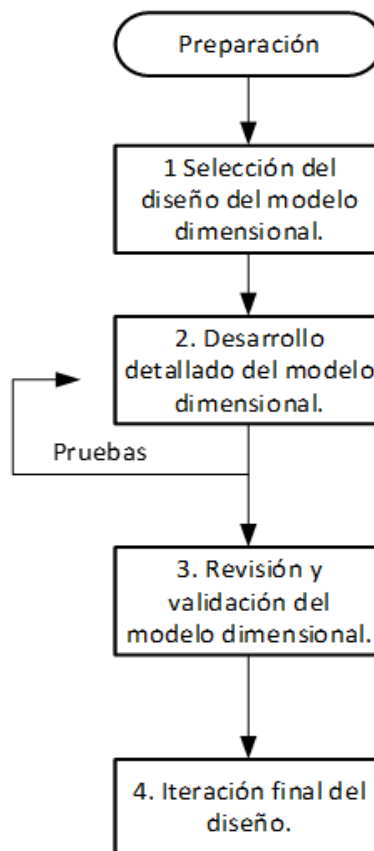


Figura 3. Diagrama de flujo del proceso dimensional de la metodología de Kimball, denominada *business dimensional lifecycle*. Fuentes: Kimball et al., 1998; Mundy et al., 2006.

El proceso de diseño empieza con un modelo dimensional de alto nivel que se realiza a partir de los procesos anticipados de la matriz descrita en el punto anterior. El proceso de pruebas consiste en cuatro pasos (Rivadera, 2010):

- ▶ Elegir el proceso de negocio: acción, tarea para ser modelada.
- ▶ Establecer el nivel de granularidad: especificar al máximo nivel de detalle.
- ▶ Elegir las dimensiones: estas nacen de las reuniones de los equipos de trabajo.
- ▶ Identificar medidas y las tablas de hechos: son los datos cuantificables del negocio que se desean analizar.

En la siguiente parte de la sesión inicial de diseño se completan las tablas con una lista de atributos bien definida. Esta se forma colocando en las filas los atributos de la tabla y en las columnas, la siguiente información: características relacionadas con la futura tabla dimensional del DW, origen de los datos y definición de los procesos ETL.

Una vez definidos los atributos y variables que van a conformar la **matriz**, se procede con la implantación del modelo dimensional. Se realizan pruebas de este y, si se presentan problemas, se hace una revisión y ajuste del modelo en caso de que sea necesario. En caso contrario, se procede a la validación del diseño.

Diseño físico

El modelado dimensional es convertido en un modelo físico y se deben tener en cuenta aspectos técnicos de *hardware* (memorias, procesadores), *software* y redes (tarjetas de red, tipos de redes, etc.). Además, hay que considerar aspectos como el tamaño del sistema *data warehouse* que se va a implementar, la configuración del sistema y las definiciones de la complejidad del modelo y del tipo de usuario que va a acceder el *data warehouse*.

Diseño del sistema de extracción, transformación y carga (ETL)

El sistema de ETL es la base sobre la cual se construye el *data warehouse*. Crear un sistema ETL que permita extraer los datos de los sistemas originales (bases de datos, tablas simples, etc.), aplicar las diferentes reglas para mejorar su calidad y consistencia, consolidar la información proveniente de distintos sistemas y, finalmente, cargar la información en el *data warehouse* en un formato adecuado (definido previamente) para la utilización por parte de las diferentes herramientas de análisis, adecuadas a las empresas (Rivadera, 2010).

Una vez finalizado el proceso ETL, se procede a la implantación del *data warehouse* en la organización. Se debe destacar que este puede ir sufriendo modificaciones después de ser implantado; esto se debe a que, una vez interactúa con los usuarios,

van surgiendo mejoras o ajustes.

Para tener una base sólida sobre cómo se ha llevado a cabo cada etapa del proceso del DW, es necesario ir documentando todo el proyecto en avances cortos para construir un informe final completo que sirva como medio de consulta y, además, como estándar para futuros *data warehouse*.

4.3. Metodologías PMI

Actualmente, existen varias metodologías para la dirección de los proyectos que también son aplicables a proyectos *business intelligence* y se adaptan a las necesidades de la empresa. Las más importantes son Scrum y PMI (siglas de Project Management Institute).

El PMI es una asociación profesional de las más grandes del mundo que cuenta con medio millón de miembros e individuos titulares de sus certificaciones en 180 países (Vahos et al., 2013).

La metodología del PMI de gestión de proyectos se compone de dos elementos esenciales: la ejecución de procesos y las áreas de conocimiento.

Proceso

Para la Guía PMBOK (*A guide to the Project Management Body of Knowledge*), no puede hablarse de proyecto si este no se concibe como un proceso, es decir, una serie de actividades coordinadas e interrelacionadas entre sí que deben ejecutarse con un fin específico. No importa si son muchas o pocas las etapas que componen un proceso. Lo más importante es que este tenga tal entidad. Por supuesto, el número de etapas varía en función de las exigencias de cada caso: participantes, complejidad de las tareas y plazos de entrega, entre otros. Lo realmente decisivo en la ejecución de un proyecto es tener despejadas las etapas básicas que lo conforman, que para el PMI son, según la PMBOK:

- ▶ Inicialización.
- ▶ Planificación.
- ▶ Ejecución.
- ▶ Control.

- Cierre.

En la Figura 4 se observa la interacción entre las diferentes etapas de la metodología PMI:

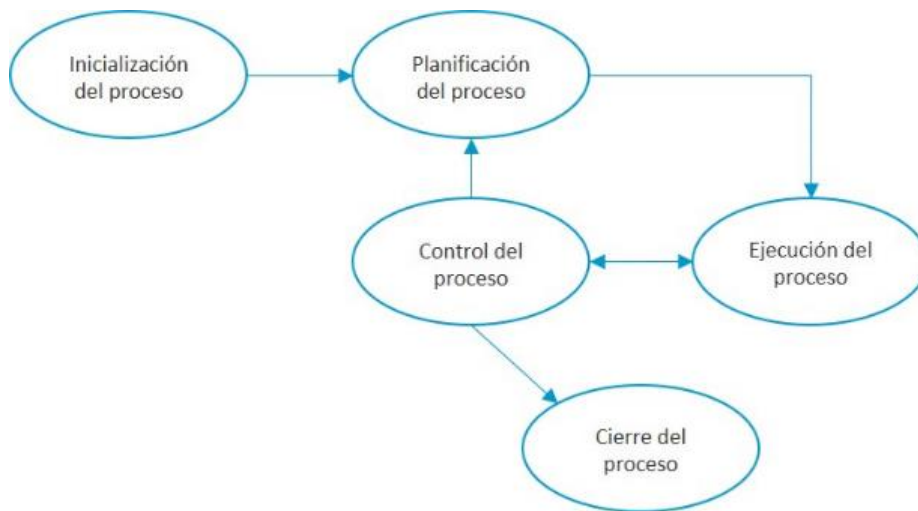


Figura 4. Etapas de la metodología PMI. Fuente: PMBOK Guide.

En el enfoque del PMI, el proceso del proyecto es realizado por personas y generalmente está en una de dos categorías:

- **Procesos de gestión de proyectos:** describe, organiza y completa el trabajo del proyecto.
- **Procesos orientados al producto:** especifica y crea el producto del proyecto, que generalmente está definido por el ciclo de vida del mismo (PMBOK).

Áreas de conocimiento

Además de los conocimientos específicos de los diferentes sectores, los líderes de proyectos deben aplicar otro tipo de conocimientos adicionales, los cuales están relacionados con competencias específicas de la gestión. Ya no es suficiente con ser un especialista en el área en la que nos desenvolvemos; hace falta adquirir

competencias transversales para realizar un mejor desempeño del trabajo. Para el PMI, las áreas de conocimiento adicionales que no pueden faltar a la hora de gestionar un proyecto son:

- ▶ Integración.
- ▶ Recursos humanos.
- ▶ Costes.
- ▶ Alcance.
- ▶ Tiempo o plazos.
- ▶ Calidad de las tareas.
- ▶ Comunicación.
- ▶ Riesgos.
- ▶ Adquisiciones del proyecto.

4.4. Metodología Inmon

Bill Inmon es considerado el padre del concepto del *data warehouse* y menciona que debe cumplir con las características que se muestran en la Figura 5.

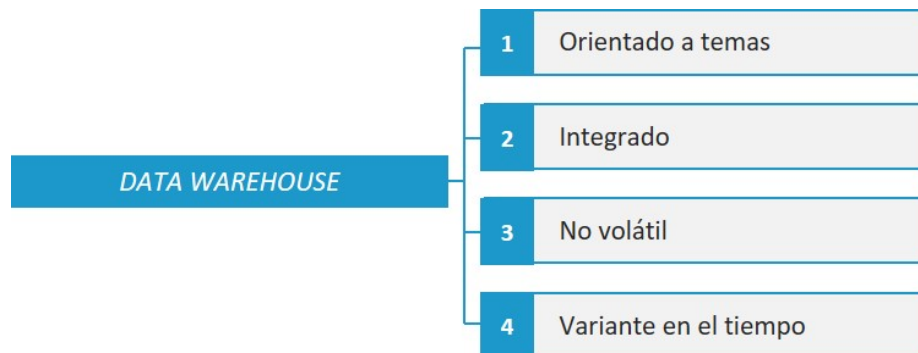


Figura 5. Características del DW según Inmon. Fuente: Inmon, 1992.

La metodología Inmon también es conocida como *top down*. Los datos son extraídos de las fuentes usando los procesos ETL y posteriormente son cargados en las áreas *stage*. Aquí son validados y agrupados en el *data warehouse* global, donde se encuentran los metadatos que contiene toda la información del *data warehouse*, es decir, los nombres de los atributos, los valores máximos y mínimos y los datos eliminados, entre otros. Una vez realizado este proceso, se actualizan los *data mart* con las respectivas transformaciones. Tener una orientación global suele ser más difícil de desarrollar e implementar.

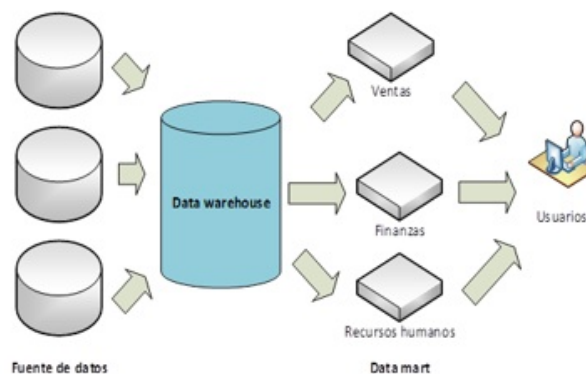


Figura 6. Metodología Inmon. Fuente: Inmon, 2002.

Ciclo de vida Inmon

El ciclo de vida para el desarrollo del *data warehouse* comienza con los datos de las diferentes fuentes, los cuales están integrados y probados. Posteriormente, se desarrollan herramientas para leer dichos datos y, por último, se definen los requisitos para los sistemas de decisiones.

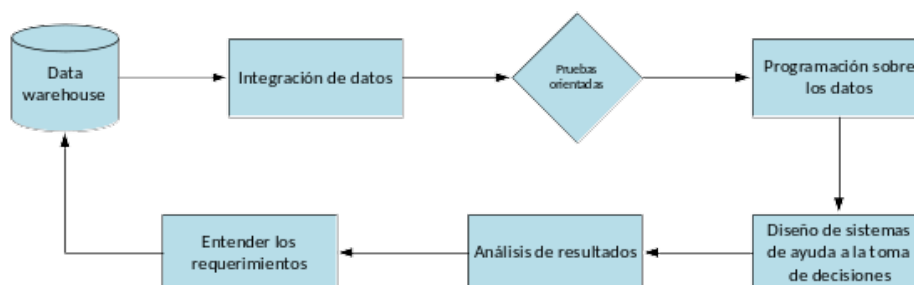


Figura 7. Ciclo de vida del *data warehouse*. Fuente: Inmon, 2002.

Estructura de un *data warehouse*

La estructura propuesta por Inmon para este sistema está dividida en cuatro niveles que se muestran en la Figura 8:



Figura 8. Estructura del *data warehouse*. Fuente: Inmon, 2002.

En la siguiente figura se muestran los cuatro niveles:

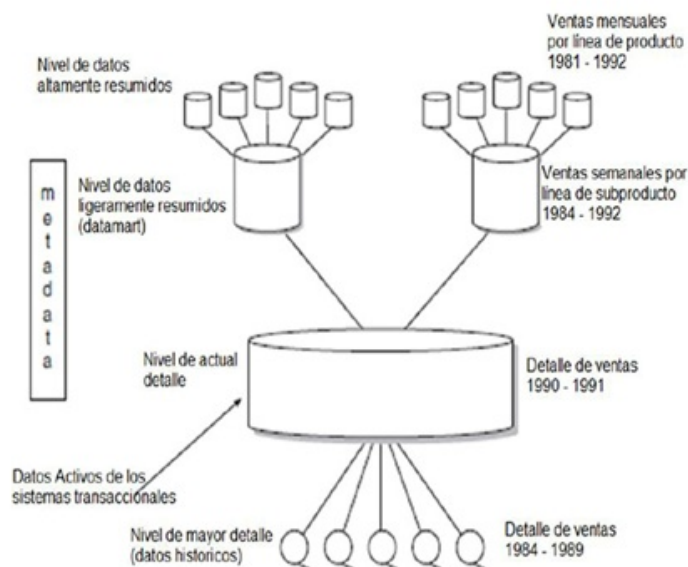


Figura 9. Estructura de los niveles del DW. Fuente: Inmon, 2002.

Proceso para la construcción del DW

Inmon propone los siguientes pasos para la construcción del DW.

1. Reconocer los sistemas transaccionales de la empresa que van a servir como fuentes de datos del DW.
2. Se empiezan a llenar las primeras tablas en el DW de las respectivas unidades del negocio. Los usuarios empiezan a acceder a los datos integrados.
3. Se cargan más tablas al DW y aumentan también los usuarios que acceden al mismo.
4. El DW es cargado con los datos correctamente, esto trae como resultado la aparición de sistemas de apoyo a la toma de decisiones.
5. Se crean los *data mart* para cada unidad de negocio.
6. Si surgen nuevas necesidades, deben crearse nuevos *data mart*.
7. Después de todos estos pasos, finalmente la arquitectura está desarrollada. En algunas ocasiones los usuarios prefieren acceder a los *data mart*, ya que su acceso es más rápido.

Granularidad

El aspecto más importante del diseño de un almacén de datos es la granularidad. De hecho, el problema de la granularidad afecta a la arquitectura que rodea el entorno del almacén de datos. La granularidad se refiere al **nivel de detalle** o **resumen de las unidades de datos en el DW**. Cuantos más detalles hay, menor es el nivel de granularidad. Cuantos menos, el nivel es mayor (Inmon, 2002).

La granularidad de los datos siempre ha sido un problema importante del diseño. Hace algunos años, la granularidad se daba por sentada. Cuando los datos detallados están siendo actualizados, es casi un hecho que los datos se almacenan en el nivel más bajo de granularidad.

La granularidad es el principal problema de diseño en el entorno del almacén de datos porque afecta profundamente el volumen de datos que reside en el DW y el tipo de consulta que puede ser respondida. El volumen de datos en un almacén se intercambia con el nivel de detalle de una consulta. En casi todos los casos, los datos llegan al almacén de datos a un nivel de granularidad demasiado alto. Esto significa que el desarrollador debe gastar muchos recursos al separar los datos.

En ocasiones, los datos ingresan al almacén a un nivel de granularidad demasiado bajo. Un ejemplo de este nivel son los datos de registro web generados por el entorno del comercio electrónico. Los datos de la secuencia de clics del registro web se deben editar, filtrar y resumir antes de que su granularidad sea adecuada para el entorno del DW (Inmon, 2002).

Los beneficios de la granularidad son:

- ▶ Después de que el almacén de datos se haya construido correctamente, proporciona a la organización una base extremadamente flexible y reutilizable para diferentes procesos de sistemas de apoyo a la toma de decisiones.
- ▶ Los datos granulares que se encuentran en el almacén de datos son la clave para la reutilización, ya que pueden ser usados por muchas personas de diferentes maneras. Por ejemplo, dentro de una empresa, los mismos datos podrían ser utilizados para satisfacer las necesidades de *marketing*, ventas y contabilidad.
- ▶ Todos estos tipos de información están estrechamente relacionados, aunque ligeramente diferentes. Con un almacén de datos, las diferentes organizaciones pueden ver los datos como lo desean.
- ▶ Visualizar los datos de diferentes maneras es solo una ventaja de tener una base sólida. Un beneficio relacionado es la capacidad de conciliar datos, si es necesario.
- ▶ Otro beneficio relacionado es la flexibilidad. Tener una base de datos sólida permite fácilmente dicha cualidad.

- ▶ Los datos granulares contienen un historial de actividades y eventos en toda la empresa. Y el nivel de granularidad debe ser lo suficientemente detallado como para que los datos puedan ser transformados.
- ▶ Tal vez el mayor beneficio de un almacenamiento de datos consiste en que los futuros requisitos desconocidos pueden ser integrados. Cuando surge un nuevo requerimiento y existe la necesidad de información, el almacén ya está disponible para el análisis y la organización está preparada para manejar los nuevos requisitos.

En la Figura 10 se muestra un ejemplo de granularidad:

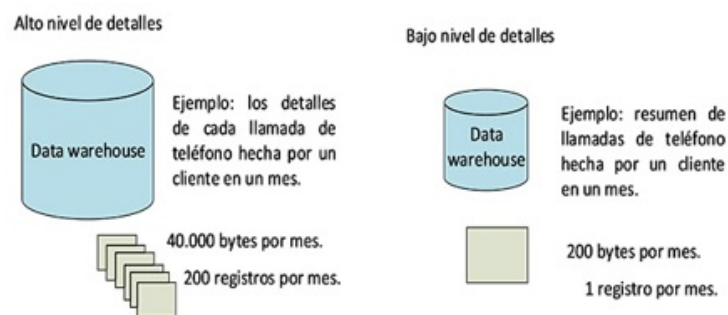


Figura 10. Ejemplo de granularidad. Fuente: Inmon, 2002.

4.5. Data-driven decision modelling

Las empresas están incluyendo el concepto de *data-driven* (literalmente, ‘impulsado por datos’) para mejorar el proceso de toma decisiones estratégicas basadas en análisis de datos y su interpretación.

Una perspectiva *data-driven* (DDM) permite a las empresas examinar y organizar sus datos con el objetivo de mejorar la atención a sus clientes y consumidores. Al usar datos para promover sus acciones, una empresa puede contextualizar y personalizar los mensajes a sus clientes o posibles clientes para un enfoque más centrado en el cliente.

Este método representa grandes avances en el modelo empírico convencional usado hasta el momento e incluye contribuciones de los siguientes campos (Solomatine et al., 2009):

- ▶ **Inteligencia artificial o AI (*artificial intelligence*):** estudio general de cómo la inteligencia humana puede ser incorporada a las computadoras.
- ▶ **Inteligencia computacional o CI (*computational intelligence*):** incluye redes neuronales, sistemas difusos y la informática evolutiva, así como otras áreas dentro de AI y máquina-aprendizaje.
- ▶ **Computación suave o SC (*soft computing*):** está cerca del CI, pero con especial énfasis en sistemas difusos basados en reglas inducidas a partir de datos.
- ▶ **Máquinas de aprendizaje automático o ML (*machine learning*):** una vez fue una subárea de AI que se concentraba en los fundamentos teóricos utilizados por CI y SC.
- ▶ **La minería de datos o DM (*data mining*) y el descubrimiento de conocimiento en bases de datos o KDD (*knowledge discovery in databases*):** se enfocan a menudo

en bases de datos muy grandes y se asocian con aplicaciones bancarias, servicios financieros y gestión de recursos de clientes. DM se ve como parte de un KDD más amplio. Los métodos utilizados provienen principalmente de estadísticas y ML.

- **Análisis inteligente de datos o IDA (*intelligent data analysis*):** tiende a centrarse en el análisis de datos en medicina e investigación e incorpora métodos de estadísticas y ML.

El DDM es un método que se centra en el uso de los métodos de CI, particularmente las máquinas de aprendizaje, para la construcción de modelos que complementan o reemplazan los modelos «basados en el conocimiento» que describen el comportamiento físico.

El proceso de construcción de un modelo DDM sigue los principios generales de: recolectar datos —seleccionar la estructura del modelo—, construir el modelo y finalmente probarlo y (posiblemente) iterar. Un enfoque general del modelo se muestra en la Figura 11.

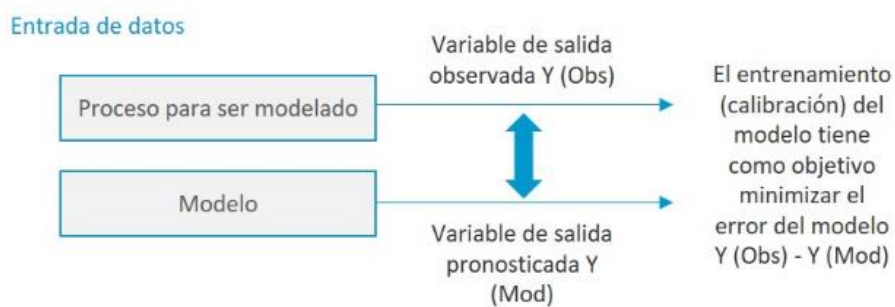


Figura 11. Enfoque general del modelo DDM. Fuente: Solomatine et al., 2009.

4.6. Metodología DevOps

Combina filosofía, prácticas y herramientas que incrementan la capacidad de la organización para desarrollar aplicaciones y servicios a gran velocidad.

Los equipos de desarrollo y operaciones ya no están aislados. Se utilizan prácticas y herramientas para automatizar los procesos que venían siendo manuales y lentos, ahora actúan de forma rápida y confiable (Microsoft, 2021).

Los beneficios de DevOps son (AWS, 2021):

- ▶ Velocidad: a través de microservicios y entrega continua.
- ▶ Entrega rápida: se automatiza el proceso de publicación de *software*, desde la creación hasta la implementación.
- ▶ Confiabilidad: monitoreo y registro ayudan a ver el desempeño en tiempo real.
- ▶ Escalado: la infraestructura del código ayuda a administrar los entornos de desarrollo, prueba y producción de forma repetible y eficaz.
- ▶ Seguridad: se puede definir y supervisar.

4.7. Nuevos roles

Con el avance de las tecnologías de la información para el análisis de los datos, se hace necesaria la incorporación de nuevos roles de usuarios que se deben integrar a los equipos de trabajo para su tratamiento.

- ▶ **Director de datos CDO (*chief data officer*):** lidera la gestión de datos y analítica asociada al negocio, es el responsable de los diferentes equipos especialistas en datos.
- ▶ **Científicos de datos (*data scientists*):** son los miembros principales del equipo de ciencia de datos, de los que extraen conocimiento e información valiosa. Tienen una perspectiva general de todo el proceso y pueden resolver problemas que se presenten en ciencias de datos, la construcción de modelos analíticos y algoritmos. Combinan diversas habilidades relacionadas con las matemáticas, la estadística, la programación y la visualización, pero también deben tener destrezas comunicativas para explicar los resultados obtenidos en la organización.
- ▶ **Ciudadano científico de datos (*citizen data scientist*):** el usuario dentro de la empresa que típicamente no está principalmente formado para este rol, pero que puede extraer valor y conocimiento a través de su experiencia, explorando los datos desde las diferentes unidades de negocio. Puede ejecutar una serie de sencillas tareas analíticas utilizando herramientas especializadas en el descubrimiento de datos. Por ejemplo, en una empresa de biotecnología hacen esta labor los mismos biólogos, expertos en el tema, conocedores de los datos e inquietos por saber sobre las últimas tecnologías de acceso y análisis de datos.
- ▶ **Ingeniero de datos (*data engineer*):** es el encargado de proporcionar los datos de una manera accesible, clara y apropiada a los usuarios y al científico de datos. Es un usuario especializado en infraestructura *big data*. Desarrolla y explota técnicas, procesos, herramientas y métodos que deben servir para el desarrollo de

aplicaciones *big data*. Tiene formación y conocimiento en gestión de bases de datos, arquitecturas de clúster (segmentación), lenguajes de programación y sistemas de procesamiento de datos.

- ▶ **Administrador de datos (*data steward*):** su responsabilidad es mantener la calidad, disponibilidad y seguridad de los datos. Busca siempre mejorar el almacenamiento y presentación de los datos en toda la empresa. Tiene conocimientos de todos los procesos de negocio y, por lo tanto, sabe cómo se utilizan dentro de estos procesos.
- ▶ **Analista de datos (*business data analyst*):** forma parte de las iniciativas y proyectos de análisis de datos. Es la persona que recopila las necesidades de los usuarios de negocio y se las pasa a los científicos de los datos de forma que sea inteligible técnicamente.
- ▶ **Artistas de datos (*data artist*):** son los responsables de crear los gráficos, infografías y otras herramientas visuales para ayudar a los diferentes usuarios de la organización a comprender datos complejos.

Además de estos roles propuestos, con la entrada en vigor del Reglamento General de Protección de Datos, RGPD (o GDPR por sus siglas en inglés), ha aparecido recientemente un nuevo rol: el ***data protection officer*** (DPO), quien se encarga de asegurar que el procesamiento de los datos personales de la plantilla, clientes, proveedores y cualquier otro individuo con los que se relacione cumple con el reglamento establecido en las reglas de protección de datos (EDPS, s.f.).

4.8. Tendencias

El *business intelligence* sigue asombrando por su capacidad de transformación, adaptación y fácil manejo, con lo que consigue que todos los usuarios puedan satisfacer sus necesidades. Las tendencias en BI ayudan a entender que este jugará un papel importante en la transformación digital de las empresas.

El autor Josep Curto, en su libro *Introducción al business intelligence*, opina que antes del año 2008 el mercado de BI había logrado una importante madurez con múltiples representantes ofreciendo soluciones que cubrían la mayoría de las funcionalidades del BI (ETL, *data warehouse*, *reporting*, cuadro de mandos y OLAP).

Esta madurez originó una fuerte consolidación del mercado durante el período 2005-2009 y lo agrupó de la siguiente forma:



Figura 12. Consolidación del mercado del BI según Josep Curto. Fuente: Curto, 2012.

- **Grandes agentes:** fabricantes externos que ampliaron su catálogo de soluciones empresariales con plataformas BI. En este entorno, las marcas más destacadas son: Oracle, que adquirió Hyperion; SAP, que adquirió Business Objects; IBM, que se hizo con el control de Cognos; y otras soluciones del mercado.

- ▶ **Empresas tradicionales del mercado:** aquellas que se mantienen con un producto especializado. Por ejemplo, Information Builders o Microstrategy.
- ▶ **Empresas de áreas especializadas:** aquellas que son expertas en un campo en concreto de la inteligencia de negocio como:
 - *Data warehouse:* Teradata, Netezza, Vertica.
 - Integración de datos: Informatica, Talend.
 - Análisis visual: Datawatch.
 - Análisis dinámico y flexible: QlikView, Tableau.
- ▶ **Empresas *open source*:** aquellas que cubren todo el proceso de la inteligencia de negocio y ofrecen soluciones con TCO (*total cost ownership*). Por ejemplo, Hitachi Vantara (quien adquirió Pentaho) y Jaspersoft (Curto, 2012).

Cada año, BI & Analytics Trend Monitor de BARC brinda a los profesionales involucrados en procesos de *business intelligence* y gestión de datos una plataforma para opinar sobre las tendencias. Las opiniones son analizadas por expertos en BI de BARC, quienes complementan y redactan un informe anual. Las tendencias en los próximos años según los análisis de los reportes del 2020 están enfocadas hacia (BARC, 2021):

- ▶
 - **Avance en analítica de datos, *machine learning*:** la cantidad de casos de uso posibles es inmensa y comprende desde realizar pronósticos sobre ingresos, precios, ventas o valor para el cliente hasta prevenir cancelaciones de contratos, optimizar el tiempo de inactividad de la máquina no planificado y muchos más. Las consideraciones de sesgo en la toma de decisiones algorítmicas y los estándares éticos para tales soluciones están ganando importancia.
 - **Desarrollo ágil de BI:** los principales beneficios del desarrollo ágil son la velocidad,

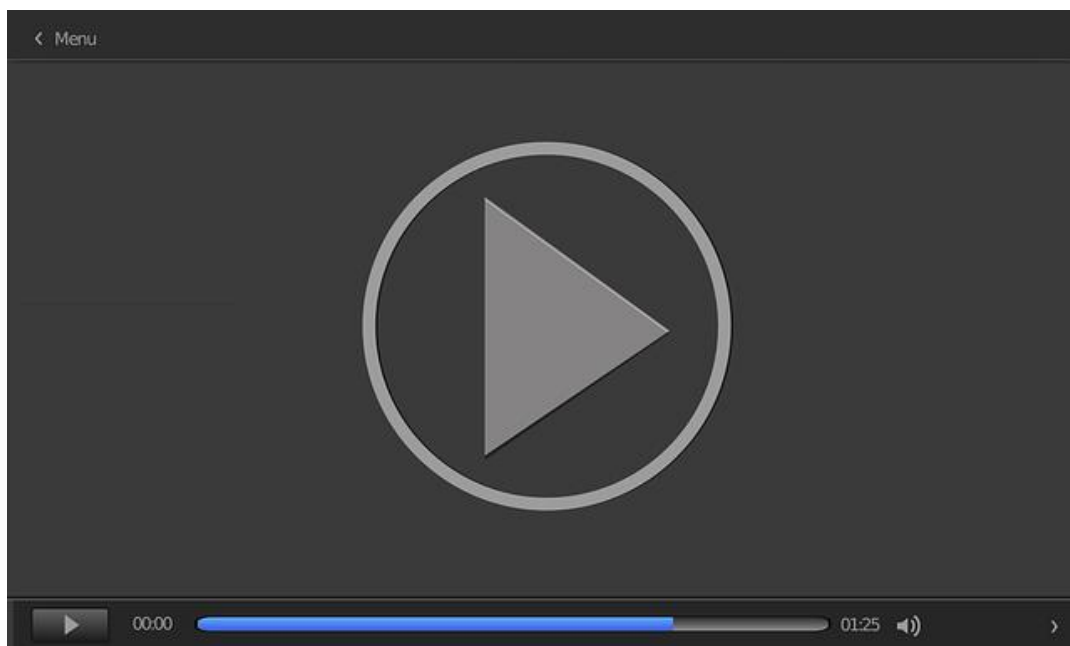
la adaptabilidad y una alineación más estrecha entre el negocio e IT. Los nuevos paneles, informes y KPI se suministran utilizando canalizaciones de datos generadas por metadatos y controladas por modelos y otros conceptos de automatización del almacén de datos. El enfoque DevOps aporta una mentalidad y las mejores prácticas para una entrega continua automatizada, que permite un cambio rápido.

- **Generación de alertas:** *machine learning* se emplea en herramientas líderes para enfocar la conciencia de los usuarios sobre tendencias y valores atípicos que antes no estaban buscando. Aquí la monetización de los datos se vuelve obvia.
- **Datos en la nube y análisis:** el reto es implementar la gestión de los datos y el análisis a la nube. Existen factores que preocupan a las empresas: factores legales, seguridad y privacidad, escasez de consejos sobre mejores prácticas para construir arquitecturas híbridas o de múltiples nubes, falta de confianza en los proveedores y el deseo de mantener bajo control los datos.
- **Maestro de datos/administración de calidad de los datos:** las decisiones correctas solo pueden tomarse con base en datos confiables, consistentes. Solo los datos maestros son fundamentales para mantener una coherencia entre informes y operaciones basadas en datos. El desafío radica en definir roles y responsabilidades para asegurar la calidad y el monitoreo continuo de los datos de una empresa.
- **Descubrimiento y visualización de datos:** descubrir patrones y datos atípicos en los datos. El aprendizaje automático guía cada vez mejor a los usuarios comerciales y automatiza tareas desde la preparación hasta la visualización. Es importante que el sistema no solo dé una respuesta, sino que proporcione explicaciones.
- **Gobierno del dato:** a diferencia de BI o BA, la gobernanza se centra en los datos de todos los sistemas que tratan con datos. Debe ser el mecanismo de dirección para la administración de los datos. Es un esfuerzo a largo plazo.
- **Preparación de datos para usuarios del negocio:** lograr una preparación de datos ágil a escala es clave para aprovechar los datos empresariales y externos y poder

tomar decisiones, automatizar procesos y monetizar los datos. Los usuarios comerciales deben contar con herramientas intuitivas y fáciles de usar que apliquen aprendizaje automático.

- **Modernización de los *data warehouse*:** IT debe estar preparado para cambio de requisitos analíticos rápidamente e implementar los DW en nuevas estructuras y más baratas a través de servicios externos.
- **Cultura *data-driven*:** cambio de las empresas que trabajan con datos aislados y orientados a proyectos a empresas completamente basadas en datos. Calidad y cantidad de datos pueden ser usados para soportar las decisiones de la empresa. No solo debe importar los beneficios de la empresa sino el nivel de cultura *data driven*.

Para finalizar, accede al vídeo Tendencias BI/BA.



Accede al vídeo:

<https://unir.cloud.panopto.eu/Panopto/Pages/Embed.aspx?id=77e855ab-4e2c-43bc-bb77-ad6600916d9b>

Vídeo. Tendencias BI/BA

4.9. Referencias bibliográficas

AWS. (2021). ¿Qué es DevOps? [Página web]. Amazon. <https://aws.amazon.com/es/devops/what-is-devops/>

BARC. (2021). Data, BI & Analytics Trend Monitor 2021 [Página web]. BARC. <http://barc-research.com/research/bi-trend-monitor/>

Curto, J. (2012). *Introducción al business intelligence*. Editorial UOC.

EDPS. (s.f.). Data Protection Officer (DPO) [Página web]. *European Data Protection Supervisor*. https://edps.europa.eu/data-protection/data-protection/reference-library/data-protection-officer-dpo_en

Inmon, W. H. (1992). *Building the data warehouse*. Wiley.

Inmon, W. H. (2002). *Building the data warehouse*. Wiley.

Kimball, R., Reeves, L., Ross, M., y Thornthwaite, W. (1998). *The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying data warehouses*. Wiley.

Microsoft. (2021). ¿Qué es DevOps? [Página web]. Microsoft Azure. <https://azure.microsoft.com/es-es/overview/what-is-devops/>

Mundy, J., Thornthwaite, W., y Kimball, R. (2006). *The Microsoft data warehouse toolkit: with SQL server 2005 and the Microsoft business intelligence toolset*. Wiley.

Rivadera, G. (2010). La metodología de Kimball para el diseño de almacenes de datos (data warehouses). *Universidad Católica de Salta*, 5, 56-71.

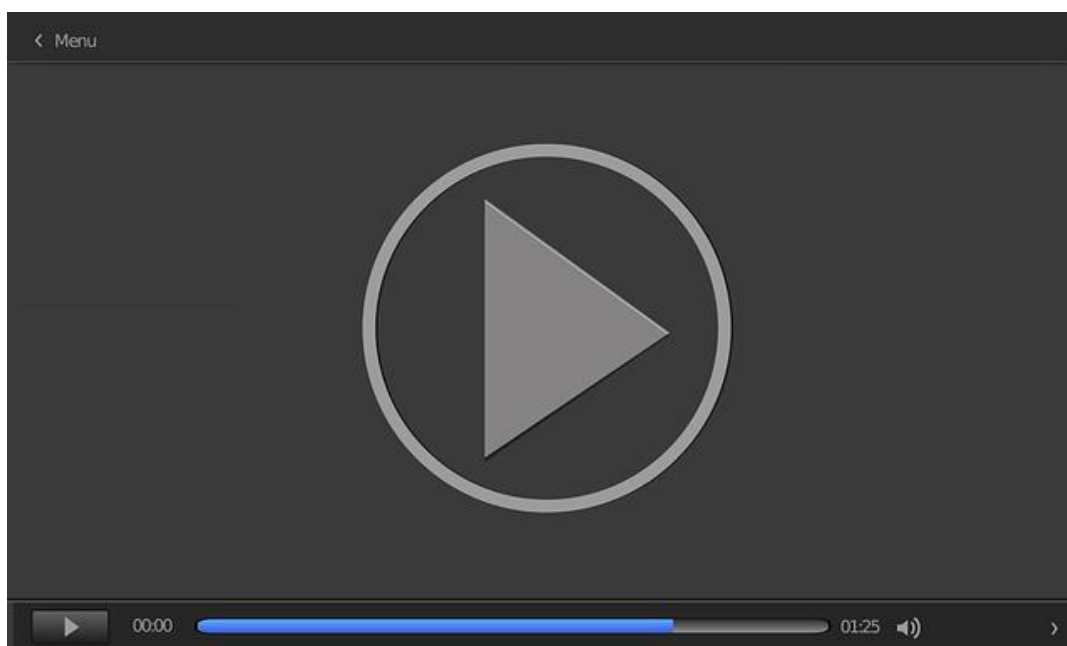
Solomatine, D., See, L. M., y Abrahart, R. J. (2009). Data-driven modelling: concepts, approaches and experiences. En *Practical Hydroinformatics* (pp. 17-30). Springer. https://www.springer.com/cda/content/document/cda_downloadaddocument/9783540798804-c1.pdf?SGWID=0-0-45-620522-p173819207

Vahos, L. E., Pastor, D. M., y Jiménez, J. A. (2013). Método para la formación de stakeholder en proyectos de ingeniería usando la metodología PMI y técnicas de inteligencia artificial. *Revista Ingenierías*, 12(23), 157-168.

Metodología de Ralph Kimball para la implementación de DW/BI

Chávez, P. (10 de mayo de 2015). *Metodología de Ralph Kimball para la Implementación de DW/BI* [Vídeo]. Youtube. <https://www.youtube.com/watch?v=f0SXElfOx6k>

Pedro Chávez nos da las claves del método Kimball, la aplicación de la inteligencia de negocio en diferentes sectores.



Accede al vídeo:

<https://www.youtube.com/embed/f0SXElfOx6k>

El concepto de DevOps

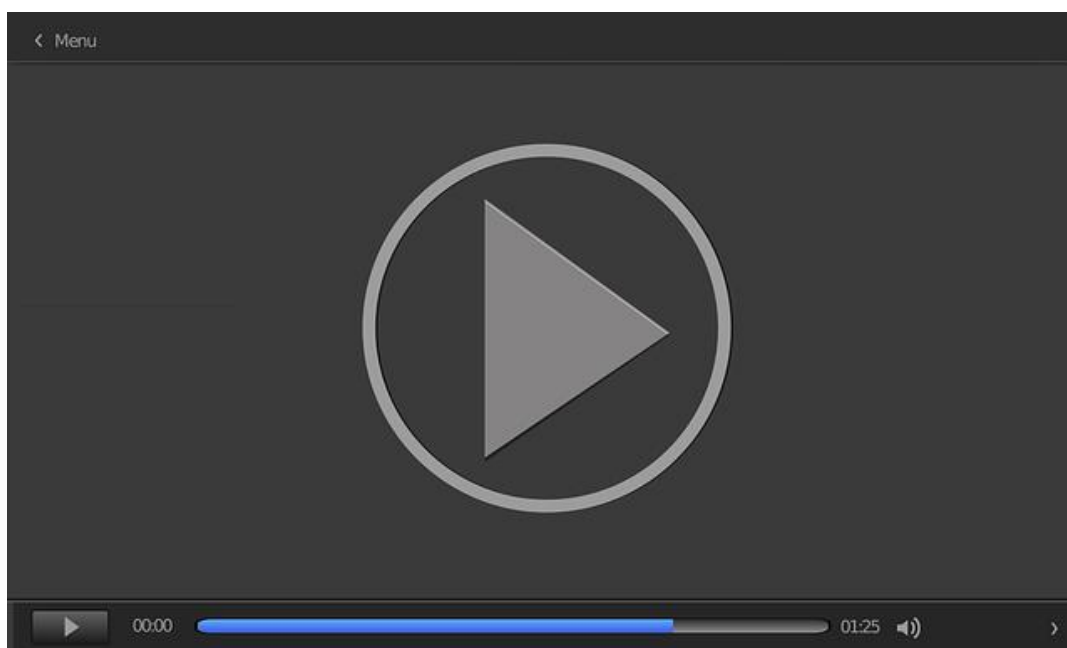
Red Hat. (2021). El concepto de DevOps [Página web]. Red Hat. <https://www.redhat.com/es/topics/devops>

Para usar DevOps se necesitan aplicaciones heredadas con las aplicaciones creadas en la nube y las nuevas infraestructuras.

Business intelligence governance best practices for 2021 and beyond

Metric Insights. (19 de marzo de 2021). *[Panel with PayPal & Veeco] Business Intelligence Governance Best Practices for 2021 and Beyond* [Vídeo]. Youtube. <https://www.youtube.com/watch?v=mcZFF0x4pws>

Abhishek Rajpurohit, Gregory Zelo y Marius Moscovici discuten en un panel sobre gobernanza y BI.



Accede al vídeo:

<https://www.youtube.com/embed/mcZFF0x4pws>

1. Son principios básicos del desarrollo de DW según Kimball:
 - A. Centrarse en el negocio.
 - B. Incrementos significativos
 - C. Los datos como centro de análisis.
 - D. A y B son correctas.

2. No es una tarea dentro de la metodología Kimball:
 - A. Crear el maestro de datos.
 - B. Diseño e implementación de ETL.
 - C. Diseño de arquitectura técnica.
 - D. Selección de herramientas de implementación.

3. Son áreas de conocimiento dentro de la metodología PMI:
 - A. DevOps.
 - B. CRISP-DM.
 - C. Costes.
 - D. Alcance.

4. ¿Cuáles son las rutas para el desarrollo del *data warehouse* según la metodología Kimball?
 - A. Tecnologías.
 - B. Procesos.
 - C. Datos.
 - D. A y C son correctas.

5. ¿Cómo se pueden obtener los requerimientos del *data warehouse*?
- A. A través de entrevistas.
 - B. Por medio de sesiones facilitadoras.
 - C. Por los clientes externos.
 - D. A y B son correctas.
6. ¿Cuántas características define Inmon para el DW?
- A. 1.
 - B. 8.
 - C. 4.
 - D. 6.
7. Son nuevos roles incorporados en los últimos años:
- A. Ingeniero de datos y analista de datos.
 - B. Consumidor de datos.
 - C. Usuarios finales.
 - D. Consumidores de información.
8. Según el estudio publicado por el BI & Analytics Trend Monitor de BARC, son tendencias del *business intelligence*:
- A. Creación de grandes empresas.
 - B. Almacenamiento de gran cantidad de datos.
 - C. Desarrollo ágil de BI.
 - D. Generación de alertas.

9. Menciona dos etapas de la metodología PMI:
- A. Pretratamiento de los datos brutos.
 - B. Inicialización y cierre de los datos maestros.
 - C. Control y ejecución del proceso.
 - D. Control y supervisión.
10. La metodología DevOps permite:
- A. Aplicar velocidad a los proyectos subidos a AWS.
 - B. Aplicar velocidad a los proyectos subidos a Azure.
 - C. Entrega rápida, confiabilidad y escalado de proyectos.
 - D. Actuar lento pero seguro en proyectos BI.