

Data Management Fundamentals

Módulo 16 – Recap CDMP & Exam Tricks

Rubén Arévalo Dosuna

GOLD SPONSORS

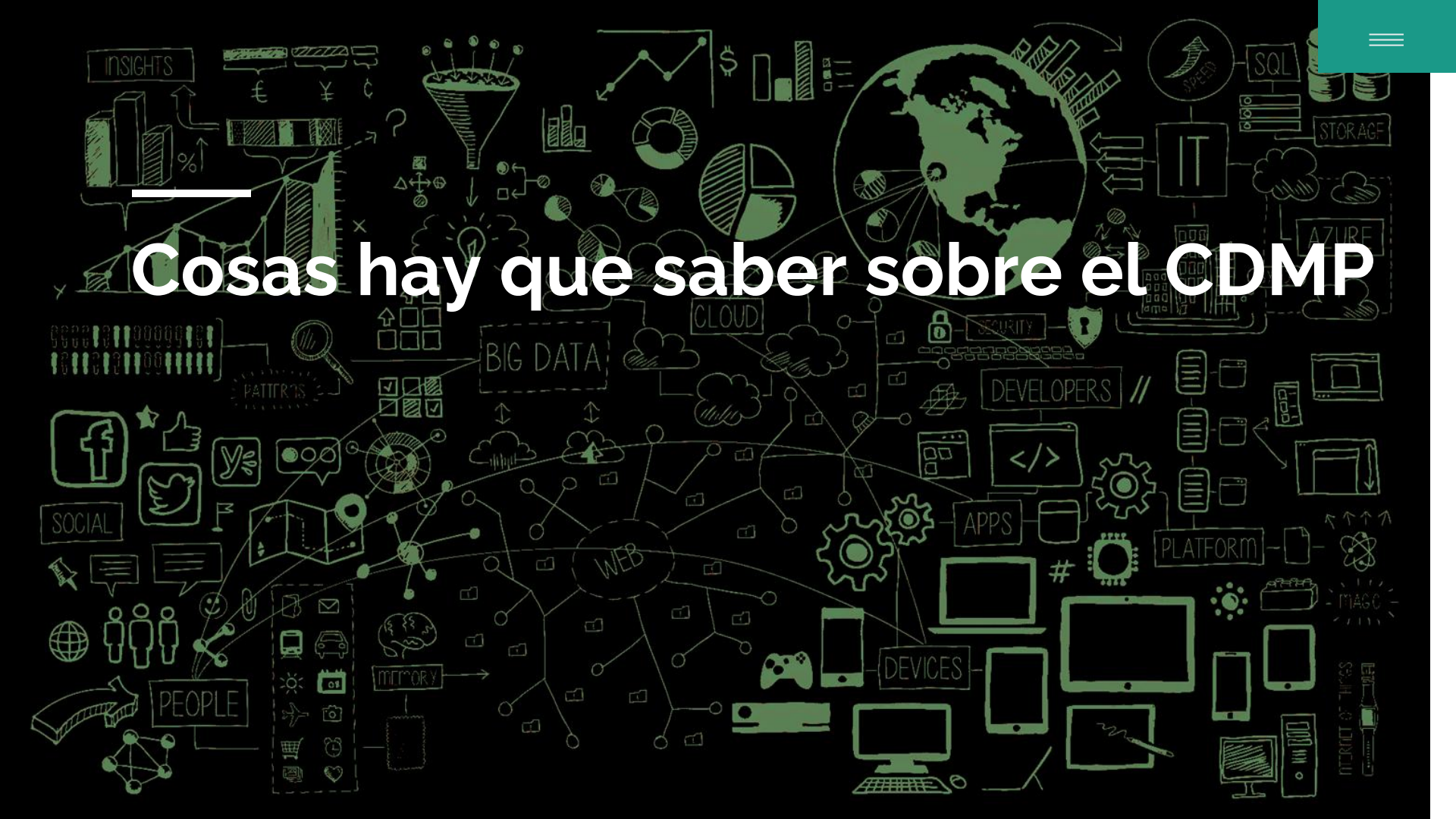


SILVER SPONSORS



Entidades académicas





Cosas hay que saber sobre el CDMP

Sobre el Examen

Todos los exámenes, independientemente del lugar donde se realicen (conferencia, formación u online), cuestan **300 USD**. Los exámenes online tienen un coste adicional de 11 dólares.

La certificación tiene una validez de 3 años. El pago del examen incluye la afiliación individual a DAMA International durante 1 año.

¿... y si fallo?

La no superación del examen le permitirá acceder a un segundo intento del mismo y tendrá un coste de **200 USD**.

No hay un periodo de tiempo mínimo entre el primer intento y la repetición.

Pay if you pass

El PayIfYouPass es un evento pensado para que todos los socios de DAMA ESPAÑA puedan presentarse GRATIS (excepto el coste de 11 dólares del registro) a los exámenes CDMP Proctored Online

Los examinados sólo tendrán que pagar en caso de obtener la certificación

Normalmente los PIYP se realizan entre los meses de Mayo y Julio

CDMP® Certification Levels And Requirements

ASSOCIATE



- > Industry experience
6 months > 5 years

- > Requirements
60% pass
Data Management Fundamentals
exam

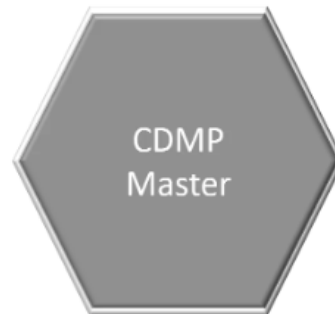
PRACTITIONER



- > Industry experience
2 years - 10 years

- > Requirements
70% pass in
Data Management Fundamentals
exam
and 70% pass in 2 specialist exams

MASTER



- > Industry experience
Minimum 10 years^

- > Requirements
80% pass in
Data Management Fundamentals
exam
and 80% pass in 2 specialist exams

Módulo 1 - Data Management Principle	2%
Módulo 2 - Big Data & Data Science	2%
Módulo 3 - Data Architecture	6%
Módulo 4 - Document and Content Management	6%
Modulo 5 - Data Ethics & Data Monetization	2%
Módulo 6 - Data Governance	11%
Módulo 7 - Data Integration and Interoperability	6%
Modulo 8 - Master and Reference Data Management	10%
Modulo 9 - Data Modelling and Design	11%
Módulo 10 - Data Quality	11%
Módulo 11 - Data Security	6%
Modulo 12 - Data Storage and Operations	6%
Modulo 13 - Data Warehousing and Business Intelligence	10%
Módulo 14 - Metadata Management	11%

100 Preguntas sobre 2000

110 minutos si optas ESL



Áreas donde te juegas el Examen

- Calidad de Datos
- Modelado y Diseño de Datos
- Datos Maestros y de Referencia
- Data Warehousing & Business Intelligence
- Metadatos



Áreas donde te juegas el Practitioner / Master

- Arquitectura de Datos
- Almacenamiento y Organización de Datos
- Seguridad de Datos
- Integración e Interoperabilidad de Datos
- Gestión de Documentos y Datos



CDMP
Certified Data
Management
Professional

Pathways

Account

Dashboard

Courses

Calendar

Inbox

Help

Welcome to the CDMP Data Fundamentals Exam. Here you can visit the study area, test your knowledge with the sample exam, take the main exam or find more information on how the exam fits into the Certified Data Management Professional (CDMP) Certification Program.



[More Information on CDMP](#)



[Study Area](#)



[Sample Exam](#)



[Main Exam](#)



Account



Dashboard



Courses



Calendar



Inbox



Help



[Home](#)

[Badges](#)

[Pathways](#)

Data Management Fundamentals Sample Exam

Due No due date

Points 40

Questions 40

Time Limit 36 Minutes

Allowed Attempts Unlimited

Instructions

Welcome! This sample exam is for you to:

1. Gauge the level of knowledge needed for the Data Management Fundamentals Exam
 2. Get used to the format and wording of the multiple choice questions
 3. Experience the speed of answering needed to finish the exam in time
- There are 40 randomly-selected multiple choice questions (5 choices for each question)
 - You have 36 minutes to complete the exam once you start
 - You have an average of 54 seconds to complete each question: the same average as the main exam
 - You can undertake a sample exam as many times as you want

[Take the Quiz](#)




Account


Dashboard


Courses


Calendar


Inbox


Help

- Home
- Badges
- Pathways

Data Management Fundamentals Exam

Started: Feb 29 at 2:32pm

Quiz Instructions

- There are 100 multiple choice questions (five choices for each question)
- When you click on the button "**Take the Quiz**", all the questions will be displayed **and** the timer will start.
- During the Exam, **a question list on the top right** allows you to see questions answered and unanswered. To navigate straight to a question, click on the required question in this list.
- After you submit, your mark will be displayed **on the right** under the heading Submission Details.

You need to link your payment details to get access to badges/certificates. Click the exam payment button for instructions

EXAM PAYMENT

Questions

- Question 1
- Question 2
- Question 3
- Question 4
- Question 5
- Question 6
- Question 7
- Question 8

Time Running: [Hide](#)
NaN Minutes, NaN Seconds

Question 1

1 pts

Knowledge area activities are classified into four categories. They are:

- ☐ Procurement, Development, Control and Operations.
- ☐ Planning, Detailing, Control and Organising.



Account



Dashboard



Courses



Calendar



Inbox



Help

☐ just-in-time reporting☐ reactive analytics☐ predictive analytics**Question 100****1 pts**

"A machine learning algorithm that incorrectly classifies new data values may have a problem with population imbalances in:"

☐ predictive analytics.☐ metadata management.☐ reference data management.☐ model training data.☐ big data architecture.

Not saved

Submit Quiz

<https://www.damaspain.org/profesionales-certificados-cdmp-por-dama-i/>

Profesionales Certificados CDMF por DAMA-I

Data Management Fundamentals

 Michelo Jurillo IBM	 Ruben Arcebal IBM	 Juana Piedra Colado Verano	 Oscar Pacheco Castillo Qualtrics	 Javier Martinez Garcia Bentley & Bentley
 Alfonso Fernandez Revinga Stratix	 Alberto Tangarino EDS	 Lucia Engo Borema Argenta Data	 Gorka Santos Ortiz NTT Data	 Alfredo Caceres Gallo Ave Networks
 Juan Males Servita	 Tomas Santamaria Tango Bank	 Jose Maria Gonzalez Serrato Oracle Software	 Nerea Sevilla Lambert	 Oscar Alonso IBM Consulting
 Francisco Huerta Sánchez CEPSA	 Alberto Rodriguez Benedict Juventi Consultores	 David Centeno Garcia B	 Ned Luzano Santamaria	 Nina Arret Santos NTT Data

 Elena Pomales Pérez NTT Data	 Pablo Ortega García Bentley	 Eduard Caceres Pérez IBM	 Rafael Nieto Ortega Telefonica Tech	 Eduard Bonichon Salas Telefonica
 Enrique Callego HMC	 Daniel Ruiz Pérez IBM	 Javier Torrealba Verano	 Ana Narta Kocika	 Belen Muñoz Villanueva Hindenburg
 Joaquín Rodríguez López Verano	 Sara Puig Aznar NTT Data	 Jose Luis Esteban Jorge	 Paul Alejandro Calacabe Tapia DM Consulting	 Manuel Blanco Carrasco DSC Technology
 Marta De Francisco Ruiz Argenta Data	 Luis Caprio Santos NTT Data	 Andrea Alabarzin HMC	 Claudia Tocchini Zurich	 Pablo Alvarez Rodriguez Kempster
 Laura Gonzalez Blanco de España	 Marc Borlani Quint	 Eduard Garcia Borja HMC	 Laura Garcia Ludema Verano	 Pablo Barachina Pascual summaOK

 Victor Manuel Fernández Rastri Verano	 Xavier Nieto Roma García	 Jose Carlos Borema Rubio Air Europe	 Pablo Diaz Vidal Accenture	 Daniel Rey López Verano
 Francisco Sánchez Pérez AISI	 Victor Orozco Delgado IBM	 Óscar Varela Sánchez Inditex	 Alfonso Horro Lamo de Espinosa Eudora	 Mario Salamanca Farto Covadonga
 Carlos Muñoz González Deltix	 Begoña López Rodriguez Ave	 Almudena Ancozar Marciano Bentley	 Sergio Jimenez Lozano Verano	 Carolina Ferreira Mayoral Ave
 Alejandro Cardan Accenture	 Miguel Ángel García Martos Allfunds	 Miguel Bonet López NTT Data	 Iker Beguirain	 Judith Sancho NTT Data
 Luis Gómez Aras Perquena Ruedas	 Sergio Hernández Fernández NTT Data	 Pedro Domínguez Schneider Electric	 M. Victoria Corozo Ave	 Marek Nowosielski Liberty Seguros

 Carlos A. del Álamo Alonso Warner Bros	 Alejandra González Reyes Stratix	 Juan Varga Copas	 Sofía de la Peña García Cegemini	 Carlos Romero San Pablo Cegemini
 Joan Ribes Lloqueria Sanoma	 Pilar Gamero Gullis NTT Data	 Vanesa Irmme Romero Lamas Iberia	 Antonio Pons López Soprasteria	 María Elena Alonso Miquel Mujica Montañana



Common issues and questions

– Are exams open book? How do I use my 'open book' if it is in digital form?

All exams are open book exams. The open book policy allows only one 'book'.

Students are permitted to have a text book or notes, but not both. The 'one book' may be in hardcopy or digital form, but not both.

You may use your DMBok2 Rev. Ed. as your 'one book'. If you decide to use the digital version it must be on a separate device, such as a tablet or a separate laptop. It must not be connected to the computer you are using for the exam. You cannot use a second monitor on the same computer.

This second device should not be connected to the internet while you are using it during the exam.

You should make sure that the only tab you have open during your exam is your exam. All other tabs and applications should be closed during your exam.

+ I have completed the Data Management Fundamentals exam and not received a badge or certificate?

+ I have emailed CDMP® Support and I have not had a reply. When will I get a response?

+ I paid for my exam before December 2020 and can no longer access my Canvas account?

+ Can I use my qualifications in another credential system to substitute for one of the CDMP® exams?

+ I have not received a password reset link when I requested for one. What should I do?



Access and take an exam

- + How do I access my exam?
- + How do I book a time for my exam?
- + When can I sit an exam?
- + How do I sit the exam?
- + Are there any system requirements for the computer I will take my exam on?
- + Are there any system requirements at in person events?
- + Why am I being prompted for an access code to take my exam?
- + Is there a timer visible on the exam?
- + English is not my first language. Can I use a browser translator during my exam?
- + Why can't see my exam in my exam account?
- + When does my exam expire?

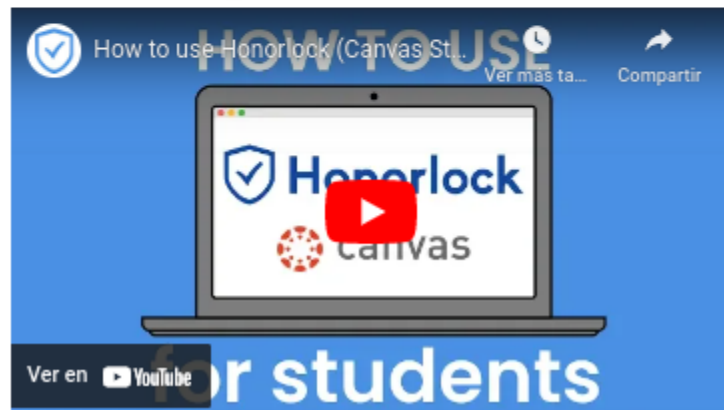


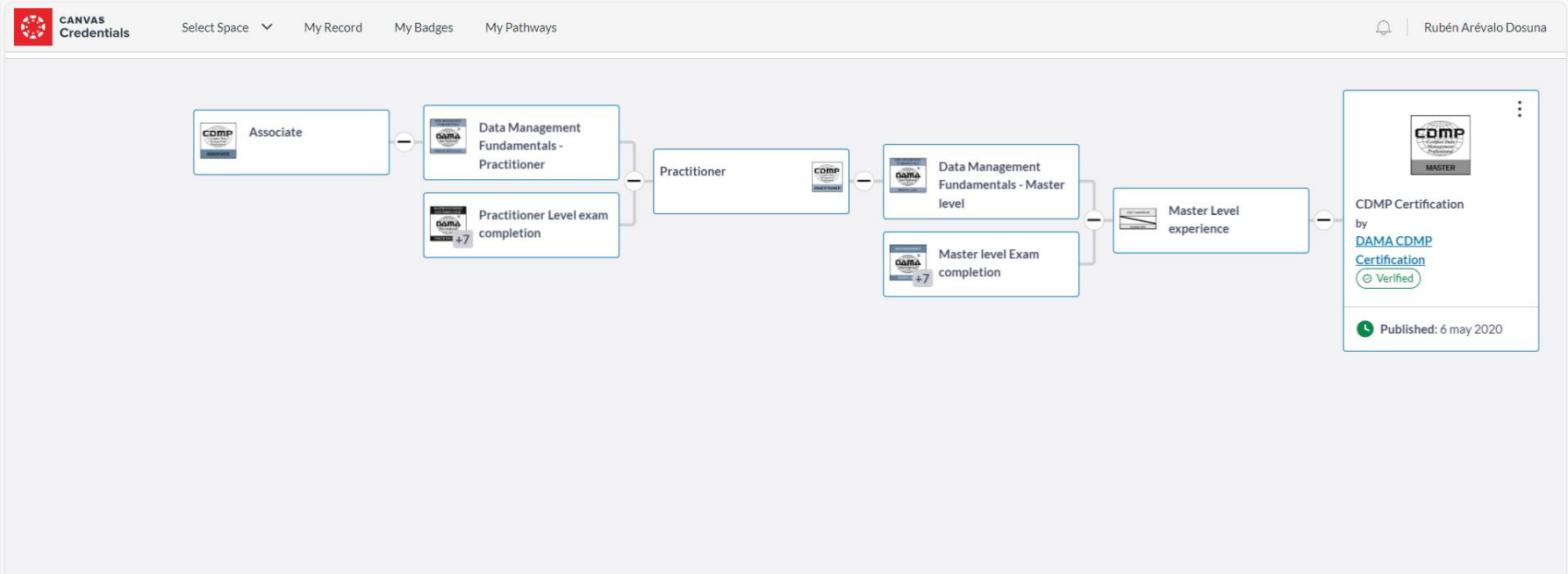
Using the online proctor

All CDMF exams taken online use Honorlock proctoring. You can find out more about how Honorlock works and how it manages privacy [here](#)

- + What do I need to prepare for the online proctor?
- + What does the online proctor do?
- + What if something goes wrong during my online exam?
- + Online proctoring how is my privacy protected?

How the verification process works





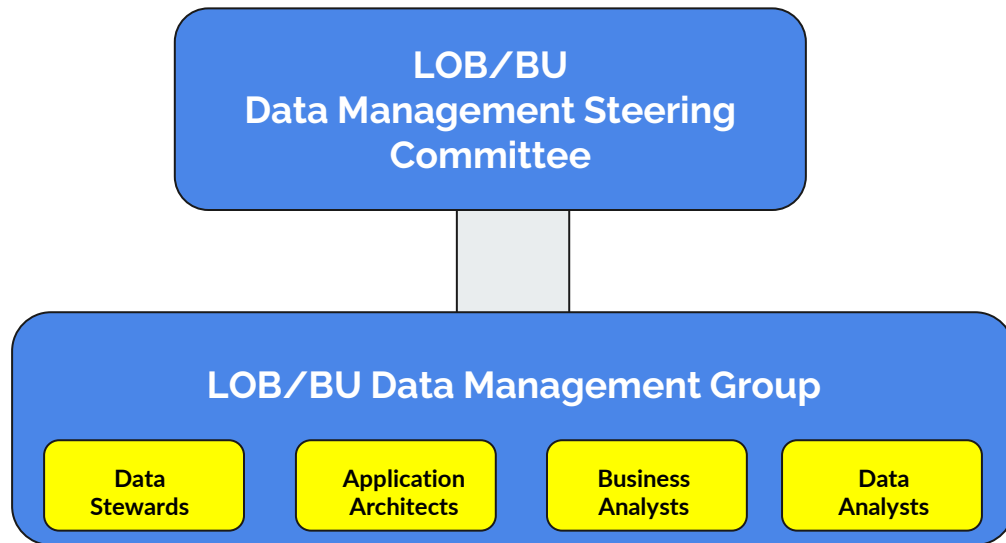


Módulo 1: Data Management Principles y Roles Expectation

Modelo operativo Descentralizado

En un modelo descentralizado, las responsabilidades de gestión de los datos se distribuyen entre diferentes líneas de negocio y TI.

La colaboración se basa en comités por LOB/BU; no hay un único propietario.



*LOB = Line of Business

*BU = Business Unit

Modelo operativo Descentralizado



Estructura relativamente plana



**Alineación de la gestión de datos
con las BUs y TI**



**Comprensión clara de las
necesidades de datos**



Fácil de implementar



**Muchos participantes en el
gobierno y la toma de decisiones**



**Decisiones colaborativas difíciles de
implementar**



**Menos formales, complicados de
mantener**



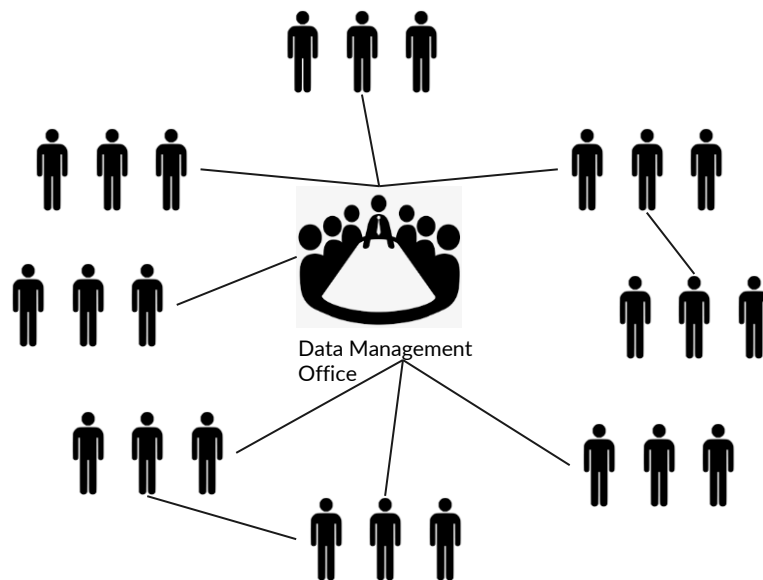
**Dificultad para definir el ownership
de los datos**



Modelo operativo de Red

La informalidad descentralizada puede hacerse más formal a través de una serie documentada de conexiones y responsabilidades a través de una matriz RACI

Funciona como una serie de conexiones conocidas entre personas y roles, y se puede diagramar como una "red"



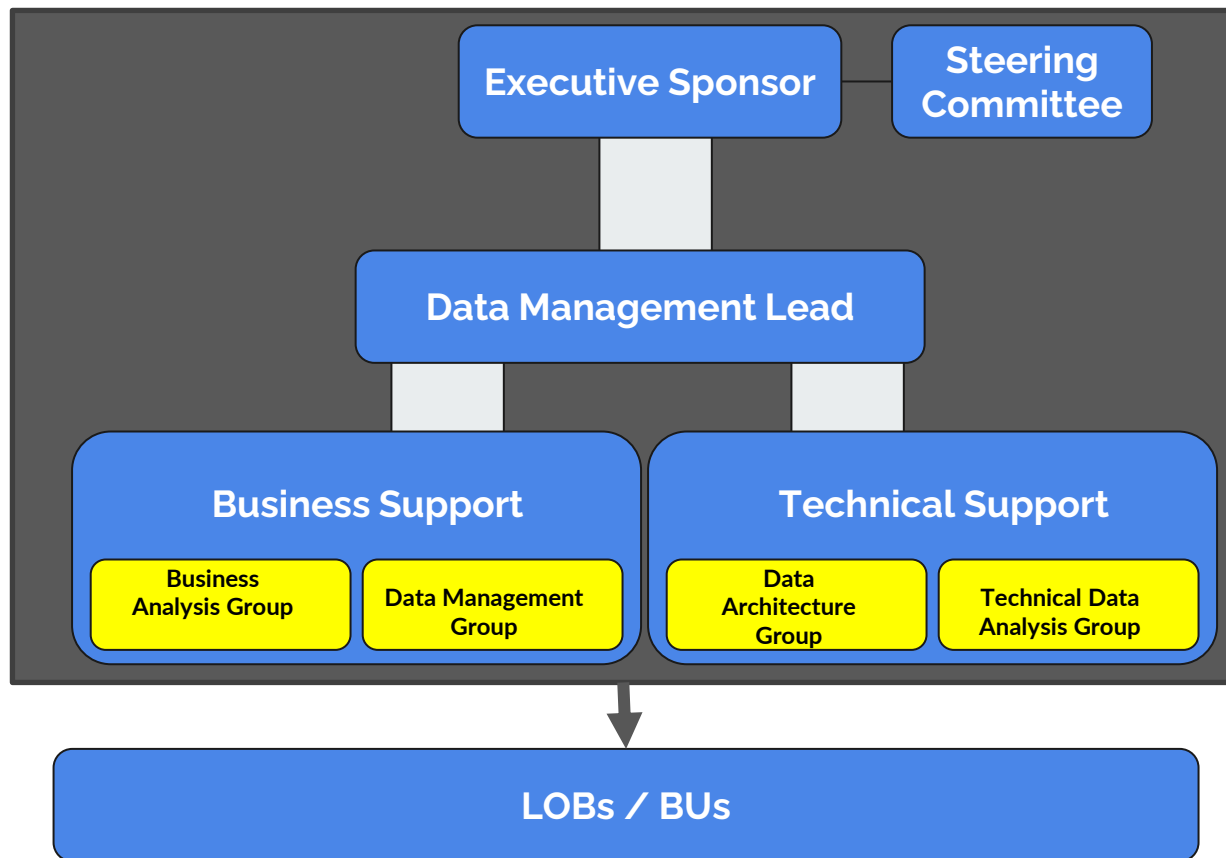
Modelo operativo de Red



Modelo operativo Centralizado

Existe un líder de gestión de datos , quien es responsable del gobierno, custodia, gestión de metadatos, gestión de datos maestros, etc...

Es el modelo operativo de gestión más formal



Modelo operativo Centralizado



Puesto ejecutivo formal



Toma de decisiones más fácil



Responsabilidad clara



Cambio organizacional significativo



Riesgo de desalineamiento con negocio



Riesgo de pérdida de conocimiento con el tiempo



Cuellos de botella en la gestión de datos en organizaciones grandes

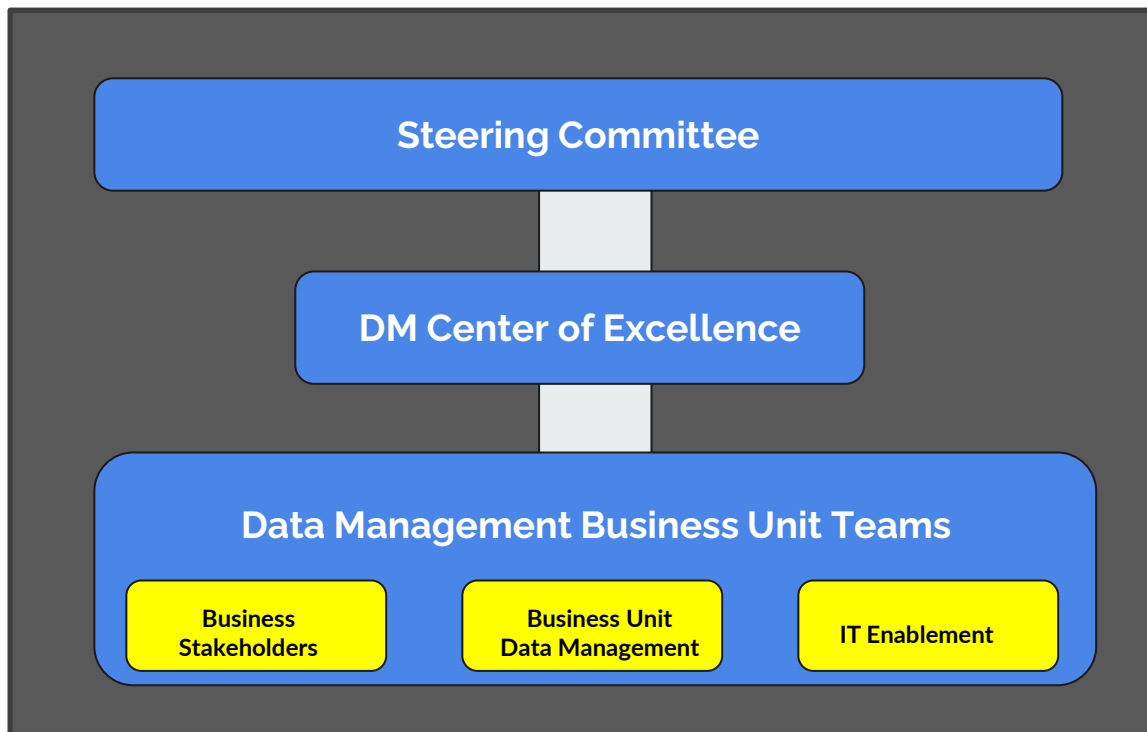
Modelo operativo Centralizado

Un modelo centralizado generalmente requiere una nueva organización. Hay que plantearse ciertas preguntas:

- ¿Dónde encaja la organización de gestión de datos en la empresa general?
- ¿Quién la dirige y a quién reporta el líder?
 - Enfoque Tecnológico CIO, Enfoque de Negocio CEO/COO
- No soporta la modalidad incremental como estrategia
- Necesita un CDO o alguien que asuma este ROL

Modelo operativo Híbrido

Abarca los beneficios de los modelos descentralizados y de los centralizados. Un Centro de Excelencia de gestión de datos centralizado trabaja con grupos de unidades de negocio descentralizadas, a través de un comité directivo ejecutivo que representa las líneas clave de los negocios y un conjunto de grupos de trabajo tácticos que abordan problemas específicos.



Modelo operativo Híbrido



Dirección adecuada desde la parte superior de la organización



Equipos de las Bus con amplia responsabilidad



Alineamiento con las prioridades de negocio



CoE dedicado para ayudar a enfocarse en desafíos específicos



Establecimiento de la organización



Personal adicional para crear el CoE



BUs con diferentes prioridades



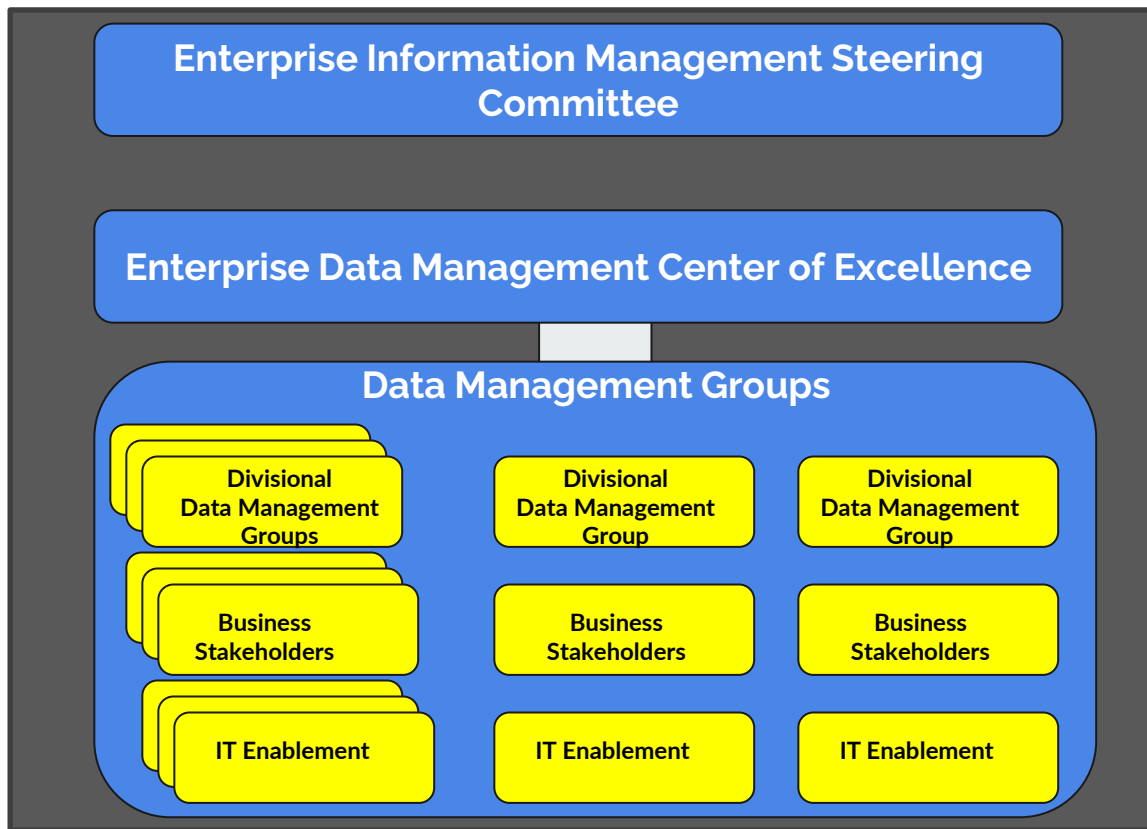
Conflictos entre el CoE y las prioridades de las organizaciones descentralizadas



Modelo operativo Federado

Es una variante del modelo operativo híbrido que proporciona capas adicionales de centralización/descentralización, que a menudo se requieren en grandes empresas globales.

Proporciona una estrategia centralizada con ejecución descentralizada.



Modelo operativo Federado



**Modelo ideal para grandes
empresas**

**Un ejecutivo de DM rinde cuentas
en toda la organización y lidera el
CoE empresarial**

**Bus facultadas para cumplir con los
requerimientos en función de
necesidades y prioridades**

**Se permite priorizar en función de
entidades de datos específicas**



Complejidad



**Equilibrio entre la autonomía de las
BUs y las necesidades de la
empresa**

Cómo diseñar un modelo organizativo de gestión de datos

Cualquiera que sea el modelo elegido, son clave la simplicidad y la facilidad de uso para la aceptación y la sostenibilidad. Si el modelo operativo se ajusta a la cultura de una empresa, entonces la gestión de datos y el gobierno adecuado pueden integrarse en las operaciones y alinearse con la estrategia. **Consejos para construir un Modelo Operativo:**

- Determinar el punto de partida evaluando el estado actual
- Vincular el modelo operativo a la estructura de la organización
- Tener en cuenta: Complejidad de la organización + Madurez // Complejidad del dominio + madurez // Escalabilidad
- Obtener el patrocinio ejecutivo - una necesidad para un modelo sostenible
- Asegurar que cualquier foro de liderazgo (comité de dirección, consejo asesor, junta) es un órgano de toma de decisiones
- Considere los programas piloto y las etapas de implementación
- Centrarse en los dominios de datos de alto valor y alto impacto
- Usar lo que ya existe
- Nunca tome un solo enfoque para todos... (One-Size-Fits-All)

Factores Críticos de Éxito según el Framework de DAMA

DAMA en su libro DmBok 2 reconoce estos 10 factores críticos de éxito para la implementación de un modelo de gestión de datos.

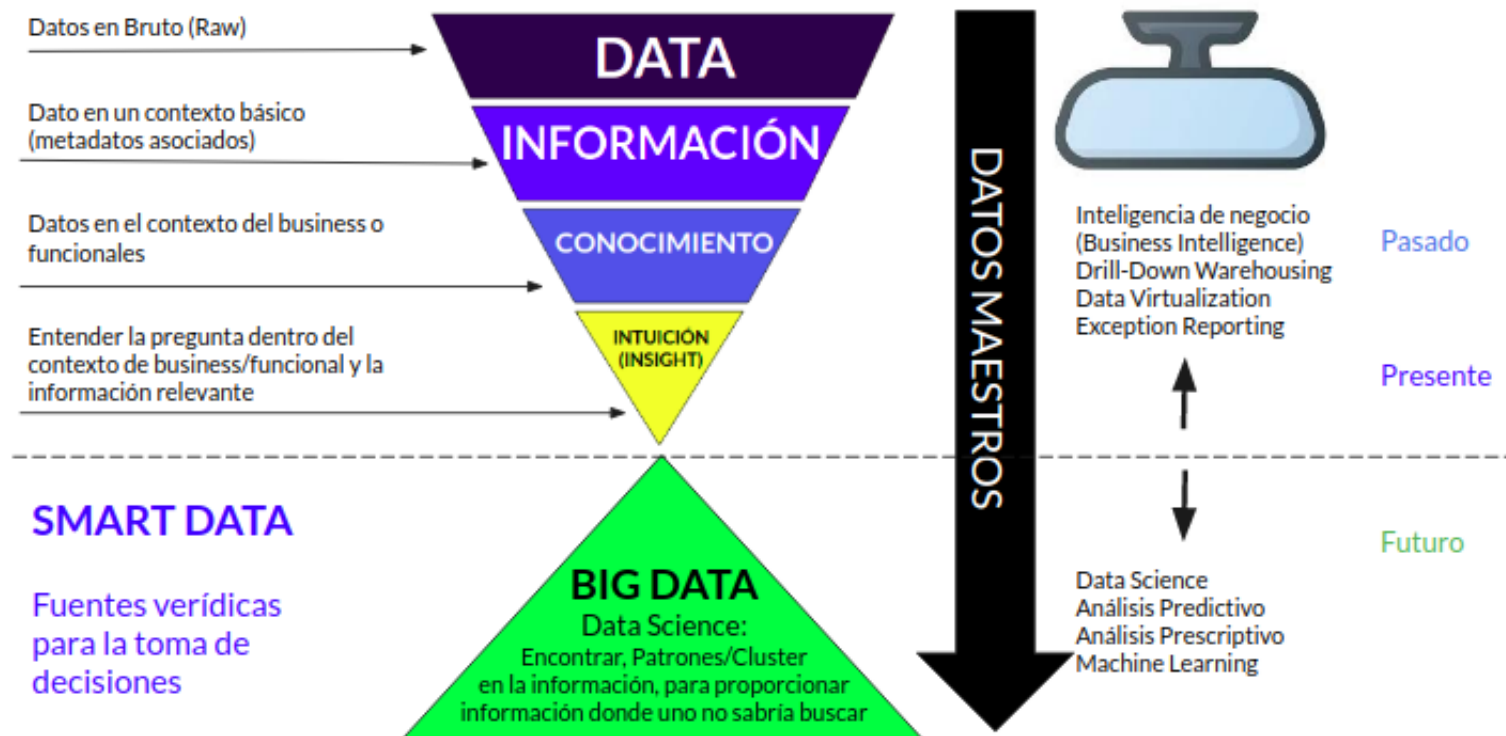
Se ha demostrado sistemáticamente que diez factores desempeñan un papel fundamental en el éxito de las organizaciones de gestión de datos eficaces, independientemente de su estructura:

1. Patrocinio ejecutivo
2. Visión clara
3. Gestión proactiva del cambio
4. Alineación del liderazgo
5. Comunicación
6. Participación de los interesados
7. Orientación y entrenamiento
8. Medición de la adopción
9. Adhesión a los principios rectores
10. Evolución, no revolución



Módulo 2: Big Data y Data Science

Abate Information Triangle^(c)



¿Qué es Big Data?

No existe una definición consensuada, sino que se basa en las V's de Big Data.

Tratamiento y acumulación de datos de todo tipo (estructurados o no) de los que se desea extraer valor.

Las 3 V's (Dough Laney, 2001)

- **Volume**, cantidad de información
- **Velocity**, velocidad de captura
- **Variety (Variability)**, diversidad de formatos y formas de captura y entrega

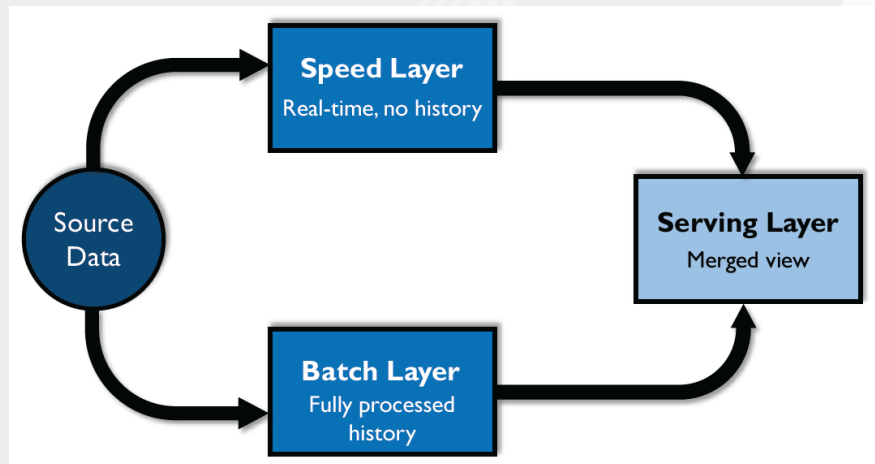
Las nuevas V's

- **Viscosity**, dificultad de uso o integración de los datos
- **Volatility**, frecuencia de cambio de los datos, cuánto tiempo son útiles
- **Veracity**, cómo de fiables son los datos

Data Lake

Entorno en el que una gran cantidad de datos de diferentes tipos y estructuras son ingestados, almacenados, evaluados y analizados.

Para procesar los datos se basa en proceso ELT en los que la transformación se ejecuta a demanda.



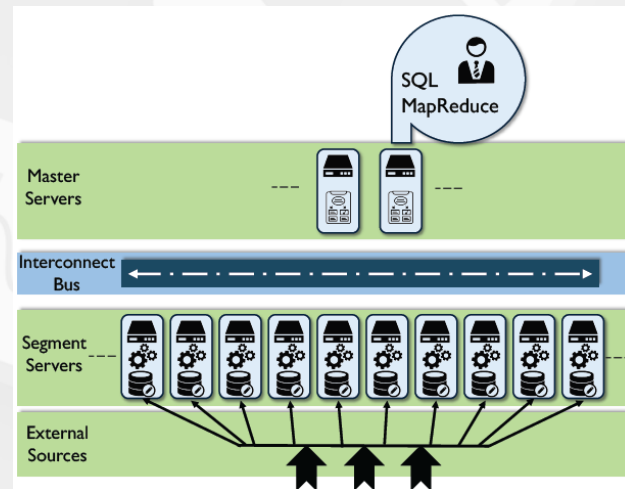
Arquitectura orientada a servicios, también se la conoce como arquitectura LAMBDA

Procesamiento masivamente paralelo (MPP)

- Los datos se distribuyen lógicamente en **múltiples servidores de procesamiento**
- **El procesamiento se distribuye** en el conjunto de servidores que trabajan con conjuntos más pequeños de datos
- **Escalabilidad lineal** a medida que se añaden nuevos servidores

Map Reduce, sistema de procesamiento de grandes conjuntos de datos. Implementado en Hadoop. Consta de 3 pasos principales que pueden ejecutarse tanto en secuencia como en paralelo:

- **Map:** Identificación y obtención de los datos
- **Shuffle:** Combinación de datos en función del patrón analítico
- **Reduce:** Eliminación de duplicados o ejecución de agregados



¿Qué es Data Science (ciencia de datos)?

La ciencia de datos **combina la minería de datos, el análisis estadístico y el aprendizaje automático** con la integración de datos y las capacidades de modelado de datos, para construir modelos de predicción que exploran los patrones de contenido de los datos.

El desarrollo de modelos predictivos a veces se denomina ciencia de datos porque el analista de datos, o científico de datos, utiliza el método científico para desarrollar y evaluar un modelo.

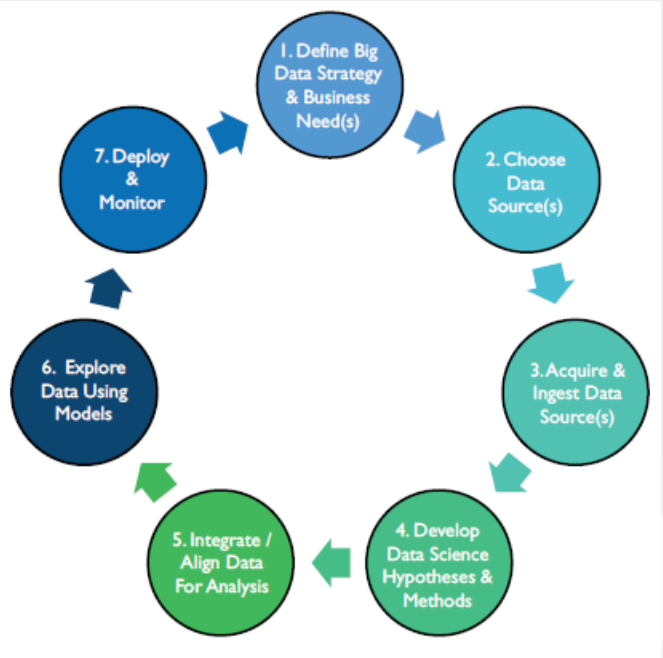
Dependencias de Data Science

El desarrollo de soluciones de Data Science implica la inclusión iterativa de fuentes de datos en modelos que generan insights. La ciencia de datos depende de:

- **Fuentes de datos ricas:** Datos con el potencial de mostrar patrones invisibles en el comportamiento de la organización o del cliente
- **Alineación y análisis de la información:** Técnicas para comprender el contenido de los datos y combinar conjuntos de datos para formular hipótesis y probar patrones significativos
- **Entrega de información:** Ejecutando modelos y algoritmos matemáticos contra los datos y produciendo visualizaciones y otros resultados para obtener una comprensión del comportamiento
- **Presentación de las conclusiones y de los datos:** Análisis y presentación de las conclusiones para que se puedan compartir los insights

El proceso de Data Science

DAMA



CRISP-DM



Big Data and Data Science

Definition:The handling of large amounts of data (Big Data) and paradigm and statistical analytics (Data Science) of many different types of data to find answers and insights.

Goals:

1. Discover relationships between data and the business.
2. Discover and analyze new factors that might affect the business.
3. Package communications of model outputs for stakeholders and decision-makers.
4. Integrate existing organizational practices with best practices in data management, big data, and data science.

Business Drivers

Inputs:

- Business Strategy
- Business Case
- Information Requirements
- Information Sources & Metadata
- IT Standards
- Analytical Models
- Data Models

Activities:

1. Define Big Data Strategy & Business Needs (P)
2. Establish Big Data Environments (D)
3. Choose Data Sources (P)
4. Acquire & Ingest Data Sources (D)
5. Develop Hypotheses & Methods (D)
6. Integrate/Align Data For Analysis (D)
7. Explore Data Using Models (D)
8. Communicate Output to Stakeholders (D)
9. Deploy and Monitor (O,C)

Deliverables:

- Big Data Strategy & Standards
- Big Data Landscape
- Data Sourcing Plan
- Acquired Data Sources
- Data Analysis and Hypotheses
- Data Insights and Findings
- Operations Plan
- Model Performance
- Model Enhancement Plan

Suppliers:

- Big Data Platform Architects
- Data Scientists
- Data Producers
- Data Suppliers
- Information Consumers

Participants:

- Data Curators
- Big Data Platform Architects
- Ingestion Architects
- Data SME's
- Data Scientists
- Analytic Design Lead
- Data Quality and Governance Specialists
- Metadata Specialists

Technical Drivers

Consumers:

- Business Partners
- Business Executives
- IT Executives
- Intermediaries
- Stakeholders
- Customers

Techniques:

- Analytic Modeling
- Machine Learning Techniques
- Big Data Modeling
- Columnar Compression
- Advanced Supervised Learning

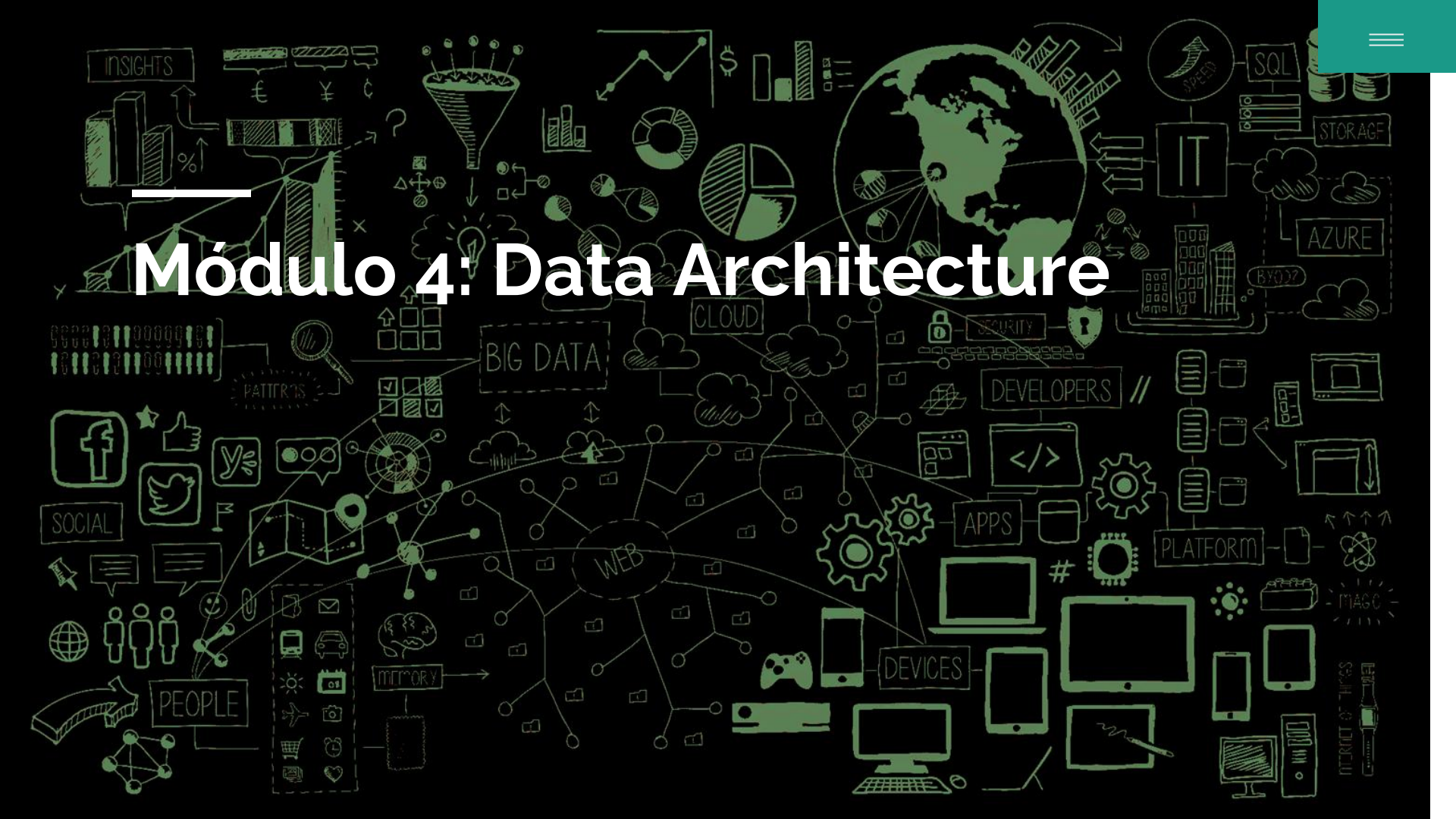
Tools:

- MPP Shared-Nothing Technologies
- Distributed File-based Databases
- In-memory Computing and Databases
- In-database Algorithms
- Big Data Cloud Solutions
- Statistical Computing and Graphical Languages
- Data Visualization Tools

Metrics:

- Technical Usage Metrics
- Loading and Scanning Metrics
- Learnings and Stories
- Response and Performance Metrics
- Data Loading and Scanning Metrics

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



Módulo 4: Data Architecture

Definición de *arquitectura*

arquitectura

Del lat. *architectūra*.

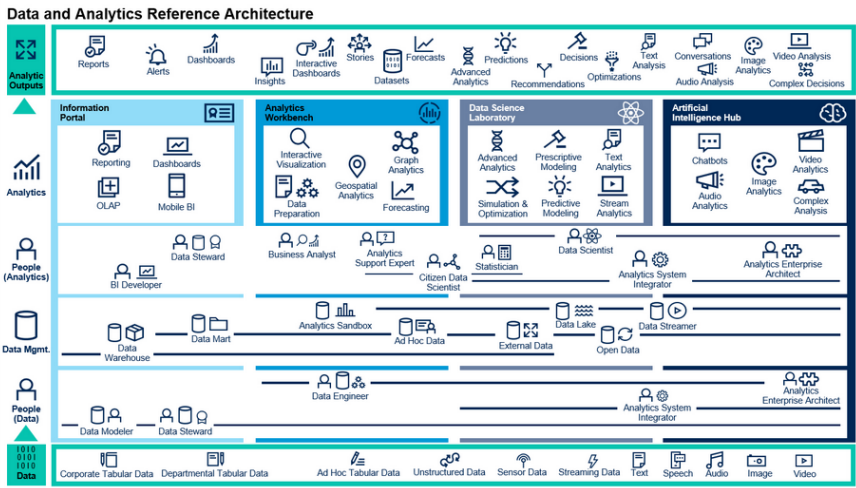
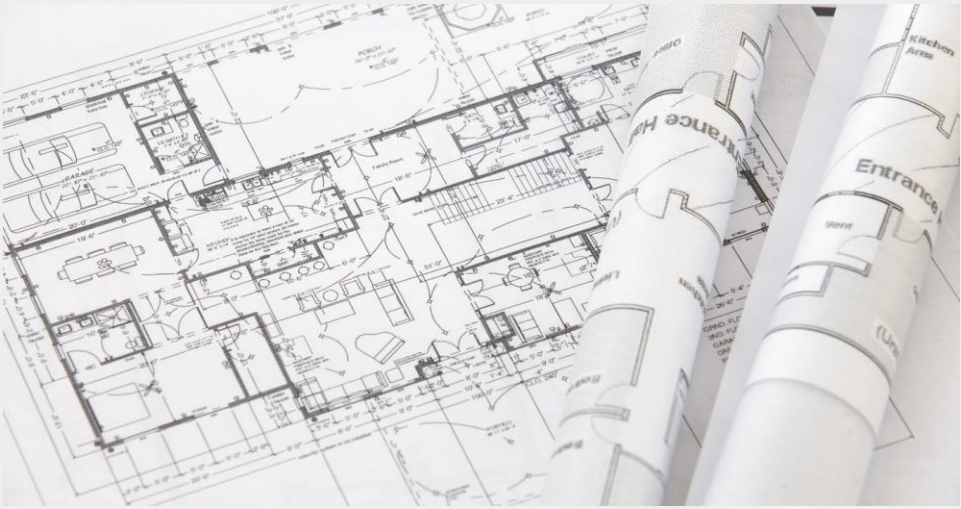
1. f. Arte de proyectar y construir edificios.
2. f. Diseño de una construcción. *Un edificio de arquitectura moderna.*
3. f. Conjunto de construcciones y edificios. *La arquitectura del centro histórico de Quito.*
4. f. *Inform.* Estructura lógica y física de los componentes de una computadora.



Systems and software engineering — Architecture description

ISO/IEC/IEEE 42010

Organización fundamental de un sistema, incorporada en sus componentes, sus relaciones entre sí y con el ambiente, y los principios que rigen su diseño y evolución



Source: Garther
ID: 463513_C

Dominios de la Arquitectura Empresarial

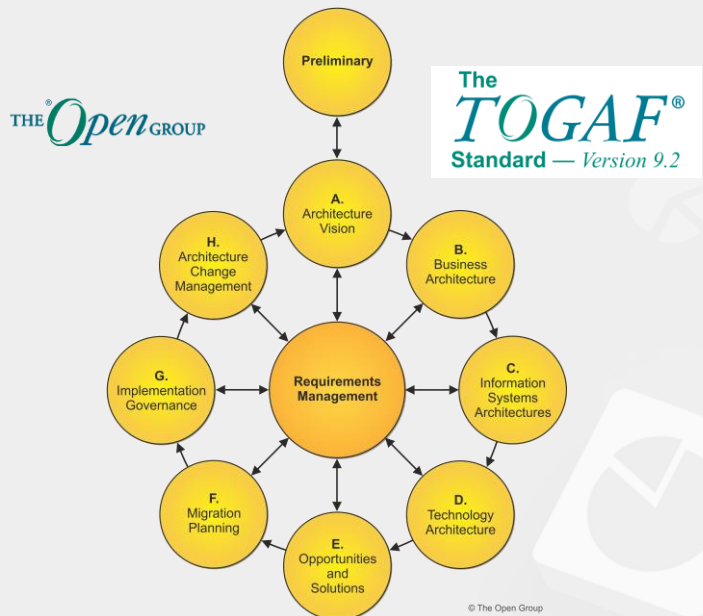
Dominio	Arquitectura de Negocio	Arquitectura de Datos	Arquitectura de Aplicaciones	Arquitectura Tecnológica
Propósito	Identificar cómo una empresa crea valor para sus clientes y otros interesados	Describir cómo los datos deben estar organizados y administrados	Describir la estructura y funcionalidad de las aplicaciones en la empresa	Describir la tecnología física necesaria para habilitar el funcionamiento de los sistemas y entregar valor
Elementos	Modelos de negocio, procesos, capacidades, servicios, eventos, estrategias, vocabulario	Modelos de datos, definiciones de datos, especificaciones de mapeo de datos, flujos de datos, APIs de datos estructurados	Sistemas de negocio, paquetes de software, bases de datos	Plataformas técnicas, redes, seguridad, herramientas de integración
Dependencias	Establece requerimientos para otros dominios	Gestiona los datos creados y requeridos por la arquitectura de negocio	Actúa en datos especificados de acuerdo con los requerimientos de negocio	Aloja y ejecuta la arquitectura de aplicación
Roles	Arquitectos de negocio y analistas, <i>data stewards</i> de negocio	Arquitectos y modeladores de datos, <i>data stewards</i>	Arquitectos de aplicación	Arquitectos de infraestructura



Systems and software engineering — Architecture description
ISO/IEC/IEEE 42010

Mantiene un estándar para la creación de Marcos (Frameworks) de arquitectura empresarial y una tabla de comparación entre los existentes

Marco de Zachman – John A. Zachman (1984, 2011)



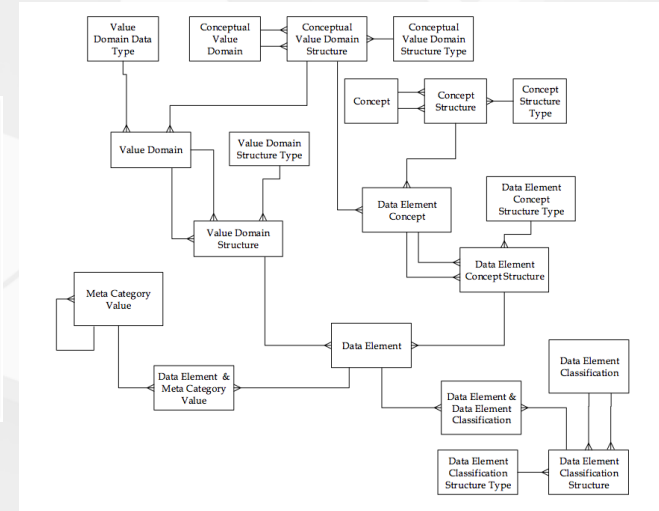
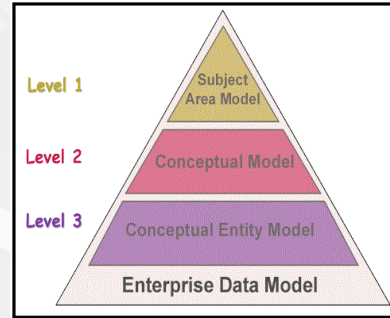
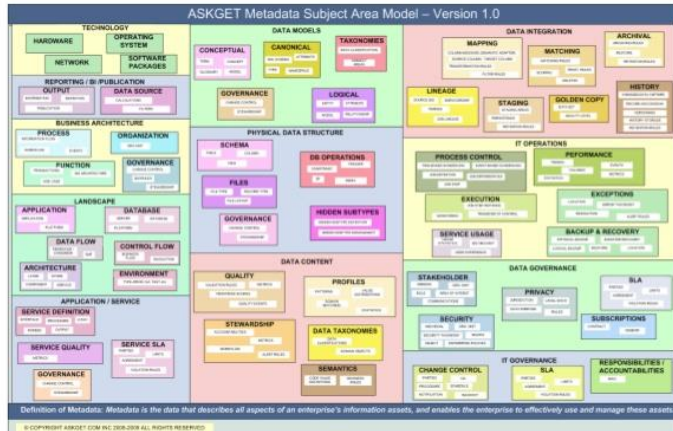
	What	How	Where	Who	When	Why	
Executive	Inventory Identification	Process Identification	Distribution Identification	Responsibility Identification	Timing Identification	Motivation Identification	Scope Context
Business Management	Inventory definition	Process Definition	Distribution Definition	Responsibility Definition	Timing Definition	Motivation Definition	Business Concepts
Architect	Inventory Representation	Process Representation	Distribution Representation	Responsibility Representation	Timing Representation	Motivation Representation	System Logic
Engineer	Inventory Specification	Process Specification	Distribution Specification	Responsibility Specification	Timing Specification	Motivation Specification	Technology Physics
Technician	Inventory Configuration	Process Configuration	Distribution Configuration	Responsibility Configuration	Timing Configuration	Motivation Configuration	Tool Components
Enterprise	Inventory Instantiations	Process Instantiations	Distribution Instantiations	Responsibility Instantiations	Timing Instantiations	Motivation Instantiations	Operational Instances
	Inventory Sets	Process Flows	Distribution Networks	Responsibility Assignments	Timing Cycles	Motivation Intentions	

Entregables de la Arquitectura de Datos

ENTREGABLES

- Diseño de Arquitectura de Datos
- Cadenas de Valor de Datos
- Modelo Empresarial de Datos
- Flujo de Datos
- Ruta de Implementación

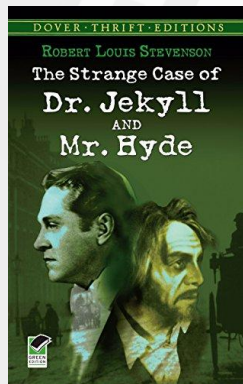
Los entregables de la Arquitectura de Datos deben reflejarse en el estado actual, el estado objetivo (arquitectura) y el estado de transición (proyecto)





Orientado a la calidad

- Se enfoca en mejorar la ejecución dentro de los ciclos de desarrollo de Negocio y TI
- Las mejoras se logran de manera incremental gracias a la involucración de los arquitectos de datos en los proyectos



Orientado a la innovación

- Se enfoca en transformar Negocio y TI para abordar nuevas expectativas
- Orientada a utilizar una lógica de negocio no probada y tecnologías de punta gracias a la colaboración con áreas de Innovación



01

Establecer la Práctica de Arquitectura de Datos

Evaluar la
Arquitectura de
Datos actual

Desarrollar una
Hoja de Ruta para
la implantación

Gestionar los
requerimientos de
los Proyectos

02

Integración con la Arquitectura Empresarial

- La Arquitectura de Datos debe ser una **parte integral** de la Arquitectura Empresarial.
- El desarrollo de especificaciones de Arquitectura de Datos se realiza típicamente dentro de **proyectos financiados** y estos generalmente impulsan las prioridades de la arquitectura.
- Los asuntos de la Arquitectura de Datos deben abordarse de **manera proactiva**, pudiendo influir en el alcance de los proyectos integrados en la gestión de Arquitectura de Datos.

- La Arquitectura de Datos describe **cómo los datos deben estar organizados y administrados** y gestiona los datos creados y requeridos por la arquitectura de negocio.
 - Los Arquitectos de Datos diseñan, mantienen y usan **planos maestros** para guiar la integración de datos, control de los activos de datos, y **alinear la inversión de datos con la estrategia de negocios**.
 - La Arquitectura de Datos **prepara estratégicamente a las organizaciones** para evolucionar rápidamente sus productos, servicios y datos para aprovechar las oportunidades comerciales inherentes a las **tecnologías emergentes**
-
- El objetivo de la Arquitectura de Datos es ser un **punto entre la estrategia empresarial y la ejecución de tecnología**.
 - La Arquitectura de Datos es más **valiosa** cuando impulsa completamente las **necesidades de toda la empresa**
 - Los **entregables** de la Arquitectura de Datos deben reflejarse en el estado actual, el estado objetivo y el estado de transición
 - La Arquitectura de Datos debe ser una **parte integral** de la Arquitectura Empresarial.



Data Architecture

Definition: Identifying the data needs of the enterprise (regardless of structure) and designing and maintaining the master blueprints to meet those needs. Using master blueprints to guide data integration, control data assets, and align data investments with business strategy.

Goals:

1. Identify data storage and processing requirements.
2. Strategically prepare organizations to quickly evolve their products, services, and data to take advantage of business opportunities inherent in emerging technologies
3. Design structures and plans to meet the current and long-term data requirements of the enterprise.

Business Drivers

Inputs:

- Enterprise Architecture
- Business Architecture
- IT Standards and Goals
- Data Strategies

Activities:

1. **Establish Enterprise Data Architecture (P)**
 1. Evaluate Existing Data Architecture Specifications
2. **Manage Enterprise Requirements within Projects (D)**
3. **Integrate with Enterprise Architecture (O)**

Deliverables:

- Data Architecture Design
- Data Flows
- Data Value Chains
- Enterprise Data Model
- Implementation Roadmap

Suppliers:

- Enterprise Architects
- Data Stewards
- Subject Matter Experts
- Data Analysts

Participants:

- Enterprise Data Architects
- Data Modelers

Consumers:

- Database Administrators
- Software Developers
- Project Managers
- Support Teams

Technical Drivers

Techniques:

- Lifecycle Reviews
- Diagramming Clarity

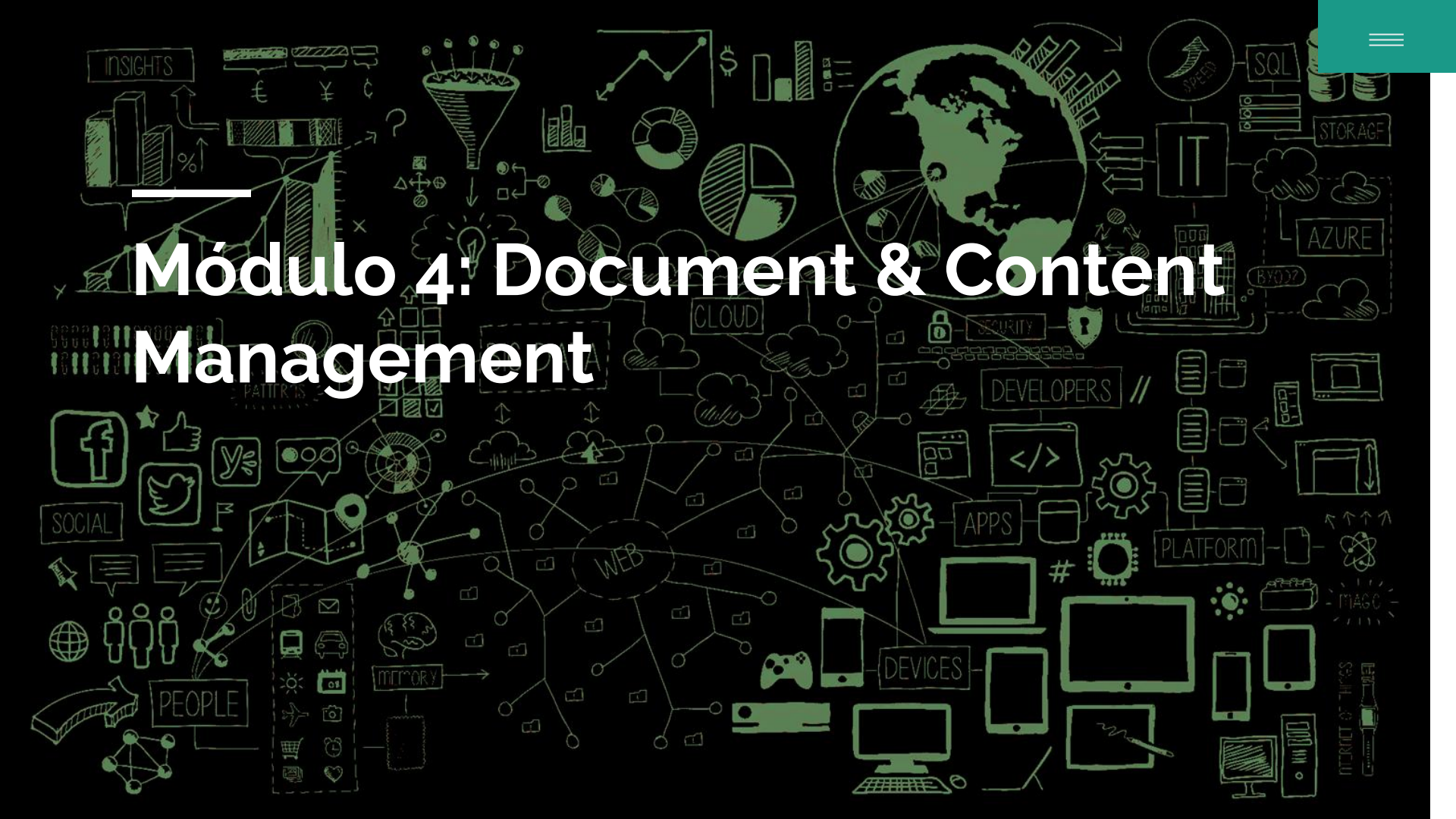
Tools:

- Data modeling tools
- Asset management software
- Graphical design applications

Metrics:

- Architecture standards compliance rates
- Trends in implementation
- Business value metrics

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



Módulo 4: Document & Content Management

Document & Content Management

- Los documentos son objetos electrónicos o en papel que se utilizan para comunicar y compartir contenidos.
- **Los registros son un subconjunto de los documentos** que proporcionan pruebas de las acciones realizadas y las decisiones tomadas en relación con el documento.
- La **Asociación de Gestores y Administradores de Registros** (ARMA) ha publicado un conjunto de Principios de Conservación de Registros Generalmente Aceptables (**GARP**) que describe cómo deben conservarse los registros empresariales. mantener los registros de las empresas.
- La gestión de documentos y contenidos (DCM) implica el **control de la captura, el almacenamiento, el acceso y el uso de datos estructurados y no estructurados almacenados fuera de las bases de datos estructuradas**. Es vital para la DCM una precisa **gestión de los metadatos**.

Document & Content Management

- Cuando el ámbito de la gestión de documentos y contenidos incluye a toda la empresa, se denomina **Gestión de Contenidos Empresariales (ECM)**. El ECM contiene la información que se muestra en los sitios web y contenida en los documentos.
- En la gestión de contenidos, una taxonomía es una estructura de nomenclatura que contiene un vocabulario controlado utilizado para delimitar temas y habilitar sistemas de navegación y búsqueda.

Document and Content Management

Definition: Planning, implementation, and control activities for lifecycle management of data and information found in any form or medium.

Goals:

1. To comply with legal obligations and customer expectations regarding Records management.
2. To ensure effective and efficient storage, retrieval, and use of Documents and Content.
3. To ensure integration capabilities between structured and unstructured Content.

Business
Drivers

Inputs:

- Business strategy
- IT strategy
- Legal retention requirements
- Text file
- Electronic format file
- Printed paper file
- Social media stream

Activities:

1. **Plan for Lifecycle Management (P)**
 1. Plan for Records Management
 2. Develop a content strategy
2. **Create Content Handling Policies, including E-discovery approach**
3. **Define Information Architecture (D)**
4. **Manage the Lifecycle (O)**
 1. Capture and Manage Records and Content (O)
 2. Retain, Dispose, and Archive Records and Content (O)
5. **Publish and Deliver Content (O)**

Deliverables:

- Content and Records Management Strategy
- Policy and procedure
- Content Repository
- Managed record in many media formats
- Audit trail and log

Suppliers:

- Legal team
- Business team
- IT team
- External party

Participants:

- Data steward
- Data management professional
- Records management staff
- Content management staff
- Web development staff
- Librarians

Consumers:

- Business user
- IT user
- Government regulatory agency
- Audit team
- External customer

Technical
Drivers

Techniques:

- Metadata tagging
- Data markup and exchange format
- Data mapping
- Storyboarding
- Infographics

Tools:

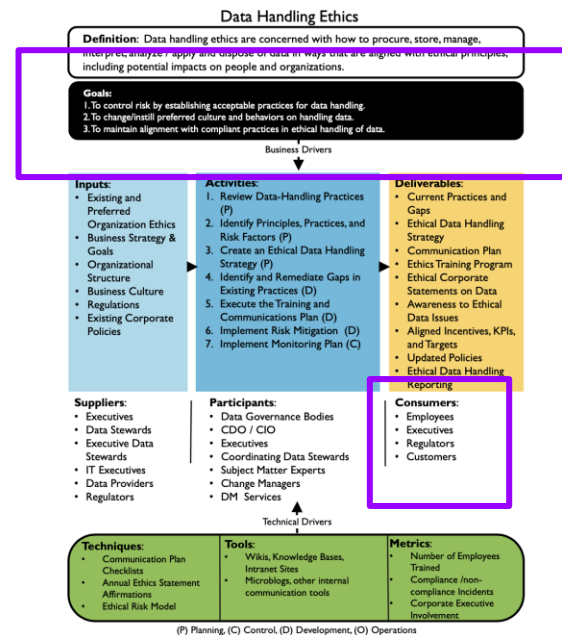
- Office productivity software
- Enterprise content management system
- Controlled vocabulary / meta-data tool
- Knowledge management wiki
- Visual media tool
- Social media
- E-discovery technology

Metrics:

- Compliance audit metric
- Return on investment
- Usage metric
- Record management KPI
- E-discovery KPI
- ECM program metric
- ECM operational metric

Objetivos de la ética en la gestión de datos

- **Controlar el riesgo**, estableciendo prácticas aceptables para el manejo de datos. (DMBOK2-R1)
- **Cambiar/influir en la cultura** y comportamientos deseados/escogidos en el manejo de datos.
- **Mantener el cumplimiento** de las prácticas de tratamiento ético de datos. (DMBOK2-R1)



Motivadores de Negocio

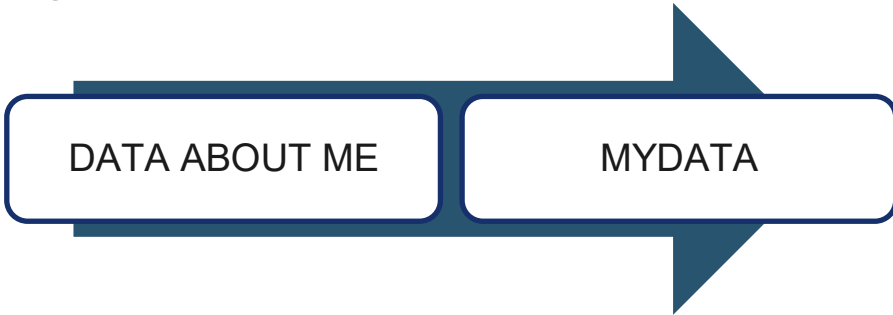
- Una de las principales razones para que una organización cultive principios éticos en la gestión de datos es la **reducción del riesgo de que los datos de los que la organización es responsable sean mal utilizados por sus empleados, clientes o socios.**



Marco normativo de referencia: GDPR

El Reglamento General de Protección de Datos (UE) 2016/679 (GDPR) es un **reglamento de la legislación de la UE** sobre la **protección de datos y la privacidad** en la Unión Europea (UE) y el Espacio Económico Europeo (EEE).

- También se ocupa de la transferencia de datos personales **fuera de la UE y el EEE**.
- El objetivo principal de la GDPR es dar **control a las personas sobre sus datos** personales y simplificar el entorno reglamentario de las empresas internacionales unificando la reglamentación dentro de la UE,



DATA ABOUT ME

MYDATA



Principio GDPR	Descripción del Principio
Justicia, Legalidad, Transparencia	Los datos personales deberán ser procesados de forma lícita, justa y en transparencia con la persona afectada.
Limitación del Propósito	Los datos personales deben ser recolectados para fines específicos, explícitos y legítimos, y no procesados de una manera que sea incompatible con esos propósitos.
Minimización de Datos	Los datos personales deben ser adecuados, pertinentes y limitados a lo que sea necesario en relación con los fines para los cuales se procesan.
Exactitud	Los datos personales deben ser exactos y cuando sea necesario, mantenerse actualizados. Se deben tomar todas las medidas razonables para garantizar que los datos personales inexactos, teniendo en cuenta el propósito para el que se procesan, sean borrados o rectificados sin demora.
Limitación de Almacenamiento	Los datos deberán conservarse en un formato que permita identificar a los sujetos [individuos] durante un período no mayor al necesario para los fines para los que se procesaron los datos personales.
Integridad y Confidencialidad	Los datos deben ser tratados de manera que garantice la seguridad adecuada de los datos personales, incluida la protección contra el procesamiento no autorizado o ilícito y contra la pérdida accidental, destrucción o daño, utilizando medidas técnicas u organizacionales apropiadas.
Responsabilidad (Accountability)	Los Controladores de Datos serán responsables y capaces de demostrar el cumplimiento de estos principios.

Bias - Sesgo

El sesgo se refiere a una **inclinación de perspectiva**. A nivel personal, el término está asociado con **juicios no razonados o prejuicios**.

Errores de Sesgo según DAMA:

- a. Recogida de datos para resultados predefinidos
- b. Presentimiento y búsqueda
- e. Uso sesgado de los datos recogidos
- f. Metodología de muestreo sesgada

Los componentes clave de una estrategia de gestión ética de los datos

Declaraciones de
valores

Principios éticos de
gestión de los datos

Marco de cumplimiento
normativo

Evaluaciones de riesgo

Formación y
comunicación

Hoja de ruta

Enfoque de la auditoría
y control

Data Handling Ethics

Definition: Data handling ethics are concerned with how to procure, store, manage, interpret, analyze / apply and dispose of data in ways that are aligned with ethical principles, including potential impacts on people and organizations.

Goals:

1. To control risk by establishing acceptable practices for data handling.
2. To change/install preferred culture and behaviors on handling data.
3. To maintain alignment with compliant practices in ethical handling of data.

Business Drivers

Inputs:

- Existing and Preferred Organization Ethics
- Business Strategy & Goals
- Organizational Structure
- Business Culture
- Regulations
- Existing Corporate Policies

Activities:

1. Review Data-Handling Practices (P)
2. Identify Principles, Practices, and Risk Factors (P)
3. Create an Ethical Data Handling Strategy (P)
4. Identify and Remediate Gaps in Existing Practices (D)
5. Execute the Training and Communications Plan (D)
6. Implement Risk Mitigation (D)
7. Implement Monitoring Plan (C)

Deliverables:

- Current Practices and Gaps
- Ethical Data Handling Strategy
- Communication Plan
- Ethics Training Program
- Ethical Corporate Statements on Data
- Awareness to Ethical Data Issues
- Aligned Incentives, KPIs, and Targets
- Updated Policies
- Ethical Data Handling Reporting

Suppliers:

- Executives
- Data Stewards
- Executive Data Stewards
- IT Executives
- Data Providers
- Regulators

Participants:

- Data Governance Bodies
- CDO / CIO
- Executives
- Coordinating Data Stewards
- Subject Matter Experts
- Change Managers
- DM Services

Consumers:

- Employees
- Executives
- Regulators
- Customers

Technical Drivers

Techniques:

- Communication Plan Checklists
- Annual Ethics Statement Affirmations
- Ethical Risk Model

Tools:

- Wikis, Knowledge Bases, Intranet Sites
- Microblogs, other internal communication tools

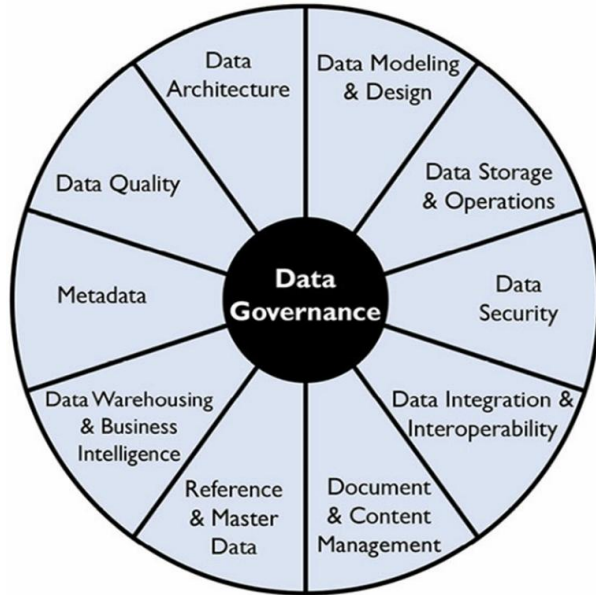
Metrics:

- Number of Employees Trained
- Compliance /non-compliance Incidents
- Corporate Executive Involvement

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

Gobierno del Dato: Capítulo 3 del DMBOK2

El Gobierno del Dato se define como el **ejercicio de autoridad y control** (planificación, monitorización y aplicación) sobre la gestión de los activos de datos.



DAMA-DMBOK2 Data Management Framework

Copyright © 2017 by DAMA International

La mayoría de funciones (**REV**) de gobierno incluyen...

- Estrategia
- Políticas
- Estándares (**DMBOK2-REV1**)
- Supervisión
- Cumplimiento
- Gestión de incidencias
- Proyectos de gestión de datos
- Valoración de activos de datos

REQUIERE UN CAMBIO CULTURAL

- El Gobierno del Dato requiere de apoyo de la **gestión del cambio organizativo** así como esponsorización por parte de **directivos C-level**

Objetivos

Definición: El ejercicio de autoridad, control y compartir la toma de decisiones (planificación, monitoreo y aplicación) sobre la gestión de los activos de datos.

Metas:

1. Habilitar una organización para gestionar sus datos como un activo.
2. Definir, aprobar, comunicar, e implementar principios, políticas, procedimientos, métricas, herramientas y responsabilidades para la gestión de datos.
3. Monitorear y guiar el cumplimiento de la política, uso de datos, y actividades de gestión.



Para cumplir estos objetivos, la función (**REV**) de Gobierno del Dato debe ser...



SOSTENIBLE

- Es un proceso continuo
- Es una función de negocio (**REV**) que requiere un compromiso organizacional.
- Significa gestionar el cambio de una manera que sea sostenible.
- Depende del liderazgo, el patrocinio y del sentido de propiedad.



EMBEBIDO

- No es un proceso adicional.
- Las actividades del Gobierno del Dato deben incorporarse a todos los niveles de la organización.



MEDIDO

- Es necesario demostrar el impacto del Gobierno del Dato.
- Es necesario comprender el punto de partida de la organización.
- Es necesario planificar una mejora medible.

La implementación de la función de Gobierno de Datos requiere un compromiso con el cambio

PRINCIPIOS PARA UNA BASE SÓLIDA



Liderazgo visionario y comprometido y **estrategia** alineada con la estrategia de negocios de la empresa



Orientación de **negocio** e impulsado por **negocio** controlando las decisiones de TI relacionadas con datos y gobernando la interacción del negocio con los datos



Responsabilidad compartida entre los *Data Stewards* y los profesionales técnicos de Gestión de Datos



Multi-capas con distintos niveles (empresarial, local e intermedios)



Basado en un **marco de referencia** y un **modelo operativo** que defina responsabilidades e interacciones



Basado en **principios formales** y mejores prácticas que se materialicen en políticas

Entregables

1. Plan estratégico y Roadmap
2. Principios, políticas y procesos
3. Framework operativo y roles de gobierno
4. Glosario de Negocio y Cuadro de Mandos
5. Plan de gestión del cambio, comunicación y formación
6. Buenas prácticas, evaluación de madurez en gestión de datos, guías, encuesta de satisfacción (**DMBOK2-REV1**)...

Principios: Normas fundamentales que establecen la filosofía y los valores en la gestión de datos.

Políticas: Directrices específicas que traducen los principios y la intención de gestión en reglas. Describen qué hacer y qué no hacer. Debería haber relativamente pocas y establecerse breve y directamente.

Procesos: Actividades operativas que implementan las políticas y ejecutan las tareas relacionadas con la gestión de datos.

Proveedores, Participantes y Consumidores

Suppliers:

- Business Executives
- Data Stewards
- Data Owners
- Subject Matter Experts
- Maturity Assessors
- Regulators
- Enterprise Architects

Participants:

- Steering Committees
- CIO
- CDO / Chief Data Stewards
- Executive Data Stewards
- Coordinating Data Stewards
- Business Data Stewards
- Data Governance Bodies
- Compliance Team
- DM Executives
- Change Managers
- Enterprise Data Architects
- Project Management Office
- Governance Bodies
- Audit
- Data Professionals

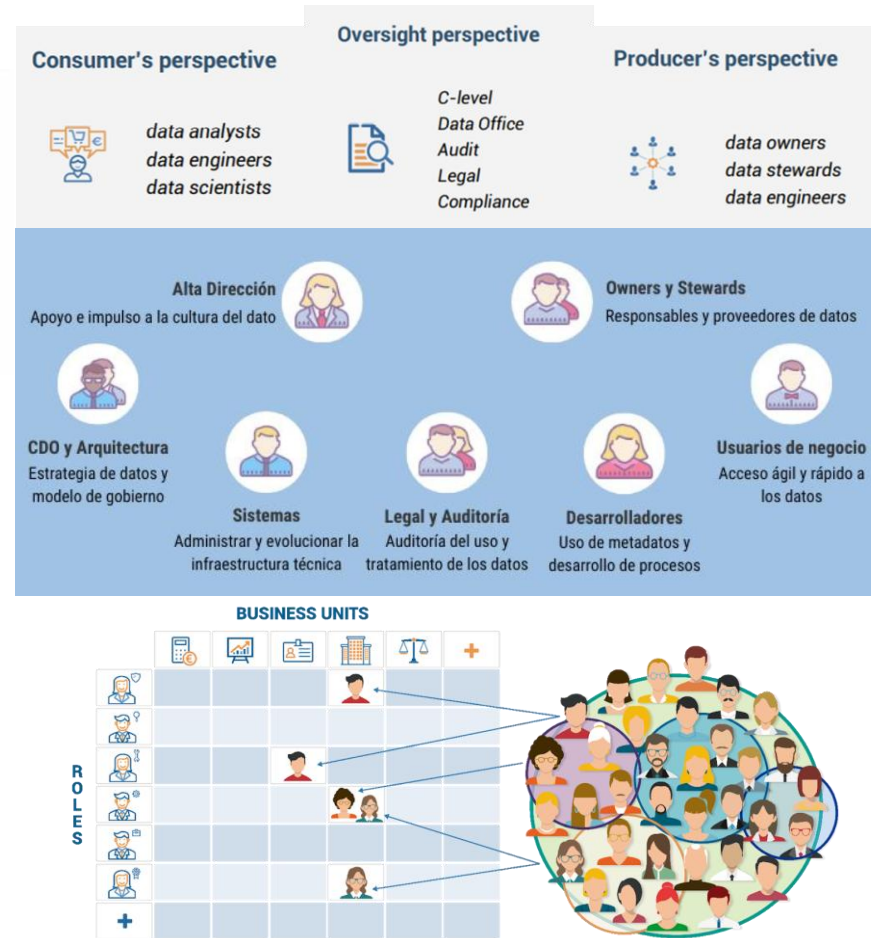
Consumers:

- Data Governance Bodies
- Project Managers
- Compliance Team
- DM Communities of Interest
- DM Team
- Business Management
- Architecture Groups
- Partner Organizations



Figuras importantes

- El Gobierno del Dato incluye **varias figuras** con funciones legislativas, judiciales y ejecutivas.
- Se suelen crear múltiples **órganos de gobierno** como Comités, Consejos, Oficinas y Grupos de Trabajo.
- El peso de los **diferentes roles** dentro del Gobierno del Dato depende en gran medida del tipo de **modelo operativo** definido e implantado.
- Las figuras más importantes son los diferentes tipos de **Data Stewards** que se pueden encontrar.



Actividades del Gobierno del Dato

01

Definir la Estrategia de Gobierno del Dato (DMBOK2-REV1)

Evaluación de la situación actual, levantamiento de necesidades y desarrollar puntos de contacto

P

02

Definir la Organización del Gobierno del Dato (DMBOK2-REV1)

Marco operativo; Objetivos, Principios y Políticas; Gestión del Cambio y de Incidencias; Cumplimiento

P

03

Implementar Gobierno del Dato

Estándares y Procedimientos; Glosario de Negocio; Coordinación con Arquitectura; Valoración de datos

O

04

Embeber el Gobierno del Dato

Incorporar, integrar y operativizar los procesos relacionados con la gestión de datos como un activo

C

O

P

PLANIFICACIÓN

C

CONTROL

O

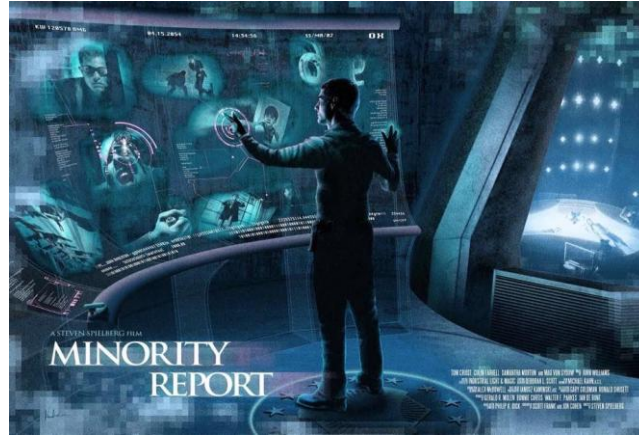
OPERACIONES



PORTALES WEB



GLOSARIO DE NEGOCIO



CUADROS DE MANDOS



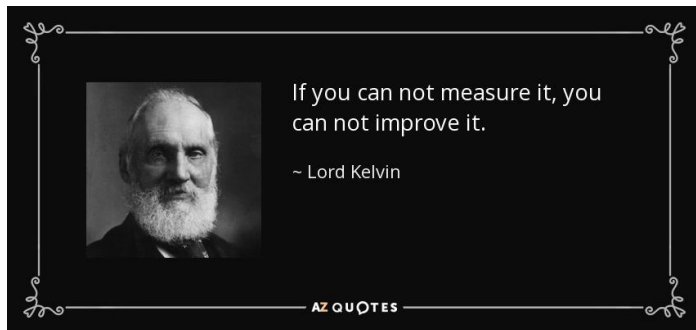
FLUJOS DE TRABAJO



GESTIÓN DOCUMENTAL

La tecnología no resuelve el problema...

- Sin embargo, existen herramientas que **ayudan** con el proceso en general.
- Es importante aprovechar los **canales de comunicación** establecidos.
- Antes de elegir una herramienta para una función específica, se deben definir los **objetivos generales** y pensar en tener un **set de herramientas**.
- Los **requerimientos funcionales** deben definirse y testearse antes de la adopción de una herramienta para no caer en el error de tener multitud de herramientas las cuales no cubren las necesidades.



Valor

- Contribución a los objetivos de la empresa
- Reducción del riesgo
- Mejora de la eficiencia de las operaciones

Eficacia

- Logro de metas y objetivos
- Uso de las herramientas
- Eficacia de la comunicación
- Eficacia de la formación
- Rapidez en la adopción de cambios

Sostenibilidad

- Desempeño de las políticas y procesos: ¿Están funcionando adecuadamente las políticas y los procedimientos definidos?
- Conformidad con las normas y procedimientos: ¿El personal está siguiendo las guías y cambiando su comportamiento según sea necesario?

Data Governance and Stewardship

Definition: The exercise of authority, control, and shared decision-making (planning, monitoring, and enforcement) over the management of data assets.

Goals:

1. Enable an organization to manage its data as an asset.
2. Define, approve, communicate, and implement principles, policies, procedures, metrics, tools, and responsibilities for data management.
3. Monitor and guide policy compliance, data usage, and management activities.

Business Drivers

Inputs:

- Business Strategies & Goals
- IT Strategies & Goals
- Data Management and Data Strategies
- Organization Policies & Standards
- Business Culture Assessment
- Data Maturity Assessment
- IT Practices
- Regulatory Requirements

Activities:

1. Define the Data Governance Strategy (P)
 1. Develop Data Governance Strategy
 2. Perform Readiness Assessment
 3. Perform Discovery and Business Alignment
 4. Develop Organizational Touchpoints
2. Define the Data Governance Organization (P)
 1. Define the Data Governance Operating Framework
 2. Develop Goals, Principles, and Policies
 3. Underwrite Data Management Projects
 4. Engage Change Management
 5. Engage in Issue Management
 6. Assess Regulatory Compliance Requirements
3. Implement Data Governance (O)
 1. Sponsor Data Standards and Procedures
 2. Develop a Business Glossary
 3. Co-ordinate with Architecture Groups
 4. Sponsor Data Asset Valuation
4. Embed Data Governance (C,O)

Deliverables:

- Data Governance Strategy
- Data Strategy
- Business / Data Governance Strategy Roadmap
- Data Principles, Data Governance Policies, Processes
- Operating Framework
- Roadmap and Implementation Strategy
- Operations Plan
- Business Glossary
- Satisfaction Survey
- Data Governance Scorecard
- Data Governance Website
- Communications Plan
- Data Valuation
- Data Management Maturity Assessment

Suppliers:

- Business Executives
- Data Stewards
- Data Owners
- Subject Matter Experts
- Maturity Assessors
- Regulators
- Enterprise Architects

Participants:

- Steering Committees
- CIO
- CDO / Chief Data Stewards
- Executive Data Stewards
- Coordinating Data Stewards
- Business Data Stewards
- Data Governance Bodies
- Compliance Team
- DM Executives
- Change Managers
- Enterprise Data Architects
- Project Management Office
- Governance Bodies
- Audit
- Data Professionals

Consumers:

- Data Governance Bodies
- Project Managers
- Compliance Team
- DM Communities of Interest
- DM Team
- Business Management
- Architecture Groups
- Partner Organizations

Technical Drivers

Techniques:

- Concise Messaging
- Contact List
- Logo

Tools:

- Websites
- Business Glossary Tools
- Workflow Tools
- Document Management Tools
- Data Governance Scorecards

Metrics:

- Compliance to regulatory and internal data policies.
- Value
- Effectiveness
- Sustainability

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



Módulo 7: Data Integration & Interoperability

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.

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



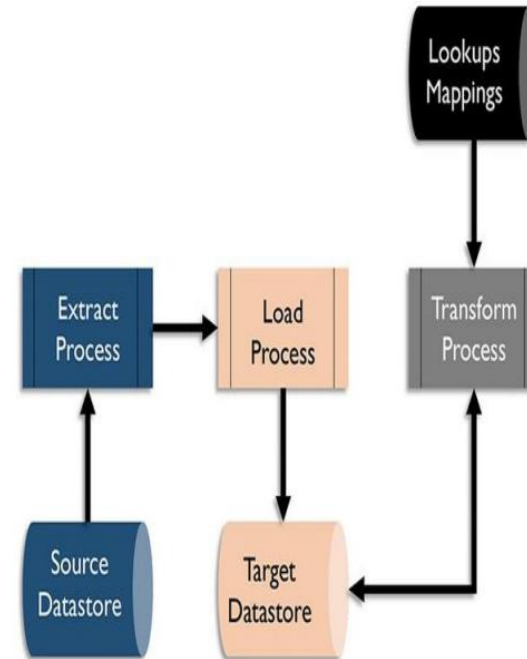
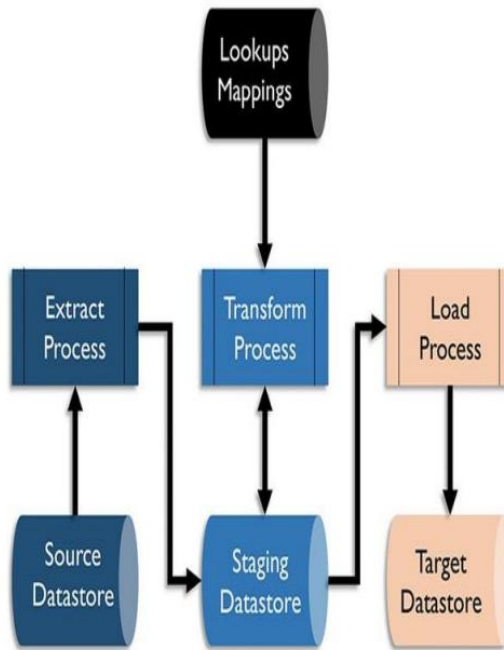
- Adoptar **una perspectiva empresarial en el diseño** para garantizar la extensibilidad futura, pero impleméntela a través de la entrega iterativa e incremental
- **Equilibrar las necesidades de datos locales con las necesidades de datos empresariales**, incluidos el soporte y el mantenimiento.
- Garantizar 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
- Latencia
- Replicación
- Archivado
- Modelo canónico
- Modelos de interacción
- Arquitecturas de integración e interoperabilidad

ETL vs ELT



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



Metadata Repository

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.

Business Drivers

Inputs:

- Business Goals & Strategies
- Data Needs & Standards
- Regulatory, Compliance, & Security Requirements
- Data, Process, Application, and Technical Architectures
- Data Semantics
- Source Data

Activities:

1. **Plan & Analyze (P)**
 1. Define data integration and lifecycle requirements
 2. Perform Data Discovery
 3. Document Data Lineage
 4. Profile Data
 5. Collect Business Rules
2. **Design DII Solutions (P)**
 1. Design DII Architecture
 2. Model Data Hubs, Interfaces, Messages, and Data Services
 3. Map Data Sources to Targets
 4. Design Data Orchestration
3. **Develop DII Solutions (D)**
 1. Develop Data Services
 2. Develop Data Flows
 3. Develop Data Migration Approach
 4. Develop a Publication Approach
 5. Develop Complex Event Processing Flows
4. **Implement and Monitor (O)**
 1. Maintain DII Metadata

Deliverables:

- DII Architecture
- Data Exchange Specifications
- Data Access Agreements
- Data Services
- Complex Event Processing Thresholds and Alerts

Suppliers:

- Data Producers
- Architecture Board
- IT Steering Committee
- Executives and Managers
- Subject Matter Experts

Participants:

- Data Architects
- Business and Data Analysts
- Data Modelers
- DII Developers
- Project and Program Managers
- Data Stewards

Consumers:

- Customers, Partners
- Information Consumers
- Knowledge Workers
- Data Scientist

Technical Drivers

Techniques:

- Loosely coupled applications
- Minimize the number of interfaces
- Create standard canonical interfaces
- Service Orchestration

Tools:

- Data Transformation Engine
- Business Rules Engine
- Data Virtualization Server
- Enterprise Service Bus
- Data and Process Modeling Tools
- Data Profiling Tool
- Metadata Repository

Metrics:

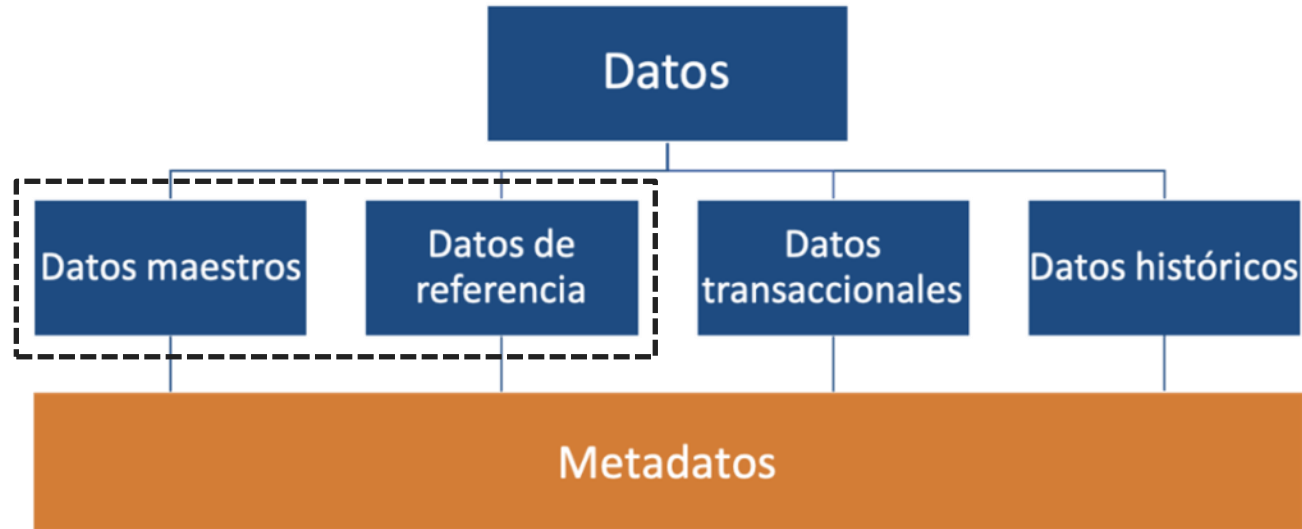
- Data Availability
- Data Volumes and Speed
- Solution Costs and Complexity

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



Módulo 8: Master y Reference Data

Conceptos esenciales – Tipos de datos



Conceptos esenciales – Tipos de datos

Clase de Datos	Amplitud de datos	Volumen	Frecuencia Actualización	Número de Columnas	Ejemplos	Propietarios o Stewards	Sistema típico de registro
Metadatos	Organización	Muy baja	Casi estática	Varias	Definiciones de atributos	Organización	Repositorio de Metadatos
Datos de Referencia Core	Organización, o casi toda la organización	Baja	Muy Baja	Pocas	SIC, Países, Códigos de Divisas, Esquemas ISSO	Normalmente externos a la organización	Repositorio de datos de referencia corporativo
Datos Maestros Comunes	Múltiples sistemas abarcando la organización	Medio	Bajo	Muchas	Cliente, Empleado, Producto, Cuentas Bancarias	Principalmente usuarios que lo crean o lo necesitan	Varios escenarios
Datos Transaccionales	Transacciones actuales	Alto	Medio	Muchas	Compras, inversiones, ...	Sistema que generaron las transiciones	Sistema que generaron las transiciones
Datos Jerárquicos	Relaciones con otros datos	Bajo	Bajo	Pocas	Jerarquías de productos, cuentas, y organizaciones	Sistema que generaron las transiciones	Sistemas de contabilidad y de productos
Datos no estructurados	Cualquier documento, correos electrónicos	Alto	Alto	Muchos	White papers, artículos, portales corporativos	Principalmente usuarios que los crean o lo necesitan	Varios escenarios
Datos de auditoría	Metadatos, datos transaccionales	Muy alto	Raramente	Pocos	Marcas de tiempo (<i>timestamp</i>), Identificadores de usuarios, log de ficheros	Organización	Sistemas que generan los datos de auditoría
Datos Derivados	Organizaciones, pero frecuentemente departamentos específicos	Muy alto	Medio	Muchos	Datamarts financieros, MDDBs	Organizacional o departamental	Sistemas que los generan

Datos de referencia - ¿Qué son?

Según el DMBOK-2 serán cualquier dato utilizado para **caracterizar o clasificar otros datos** o para **relacionar datos con información externa** a la empresa.
A lo que añadiría: de carácter NO permanente pero raramente cambiantes.

Su gestión implicara el garantizar una correcta asignación de valores en cada uno de los dominios de interés

Implica el **control de los valores** de dominio definidos y sus definiciones

Datos de referencia - Granularidad

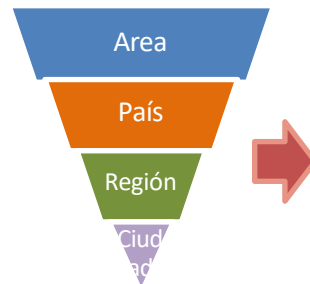
- Listas: Típica combinación código-descripción
- Listas de referencia Cruzada: lista múltiple con relación 1<->1
- Taxonomías o Jerarquías: Típicas serán las de productos o geográficas

Código	Descripción	Definición
123	Alto	Grupo de Riesgo
124	Medio	Grupo Asintomático
125	Bajo	Grupo Placebo o de Control

Ejemplo de Lista simple. . Fuente Elaboración Cesar Pecharramán para el curso Master y Reference Data.

MU	ISO	Commercial	Measurement unit text	Techn.	Measurement unit text	Dimen.
%	P1	%	%	%	Percentage	PROPOR
%O		%O	%O	%O	Per mill	PROPOR
ONE	ONE	One	One	One	One	PROPOR
D	DAY	D	Days	d	Days	TIME
225		225	mm2/s	mm2/s	Square millimeter/second	VISKIN
CMS	2M	CMS	cm/s	cm/s	Centimeter/Second	SPEED
000	2X	000	m/min	m/min	Meter/Minute	SPEED
µL	4G	µL	µl	µl	Microliter	VOLUME
µF	4O	µF	µF	µF	Microfarad	CAPACI
1B	4T	1B	pF	pF	Pikofarad	CAPACI
A	AMP	A	A	A	Ampere	ECURR
GOH	A87	GOH	GOhm	GOhm	Gigaohm	RESIST
GM3	A93	GM3	g/m3	g/m3	Gram/Cubic meter	DENSI
ACR	ACR	Acres	acre	Acres	Acres	SURFAC
AMT	MD	AMT	ADMT	ADMT	Air Dry Metric Ton	MASS

Extracto de una tabla de SAP con las unidades de Medida, mostrando una Lista de referencia cruzada. Fuente: SAP



Area	País	Region	Ciudad
EMEA	ESPAÑA	Asturias	Oviedo
EMEA	ESPAÑA	Asturias	Gijón
EMEA	ESPAÑA	Aragón	Zaragoza
EMEA	ESPAÑA	Aragón	Huesca
EMEA	ALEMANIA	Babiera	Munich
EMEA	SUDAFRICA	Región Occidental del Cabo	Ciudad del Cabo

Ejemplo de Taxonomía o Jerarquía. Fuente Elaboración Cesar Pecharramán para el curso Master y Reference Data.

Datos Maestros

DAMA los define como datos sobre las entidades de negocio que proporcionan contexto para las transacciones de negocio y análisis, en forma de conceptos comunes y abstractos que se relacionan con la actividad.

Una definición mas personal de Fernando Gualo sería: Datos comunes a varios/todos los sistemas y áreas de la organización, sobre los que pivotan las operaciones más críticas de la cadena de valor de la organización.

La gestión de Datos Maestros: implicará el asegurar los valores correctos y actuales de los mismos garantizando su interoperabilidad a través de los sistemas de la empresa.



Datos Maestros – Golden record y fuentes confiables

Una Fuente de Confianza es reconocida como la **"mejor versión de la verdad"**, basada en una combinación de reglas automatizadas y la administración manual del contenido de los datos. Una fuente confiable también puede ser referida como una Vista Única, Vista de 360°.

- Cualquier sistema MDM debe ser administrado para que sea una fuente confiable.
- Dentro de una fuente confiable, los registros que representan los datos más precisos acerca de las instancias de la entidad pueden ser referidos como Golden Records.

Gestión de datos maestros

La planificación de la gestión de datos maestros (Master data Management) incluye varios pasos básicos. Dentro de un dominio:

- **Identificar las fuentes candidatas** que proporcionarán una visión completa de las entidades de datos maestros
- **Elaborar normas** para hacer coincidir y fusionar con precisión las instancias de las entidades
- **Establecer un enfoque** para **identificar y restaurar** los datos que no coinciden y se fusionan de manera adecuada
- Establecer un **enfoque para replicar datos** confiables a los sistemas de toda la empresa

Gestión de datos maestros

El MDM es un proceso de gestión del ciclo de vida. Las actividades críticas para el ciclo de vida incluyen:

- **Establecer el contexto de las entidades de datos maestros**, incluidas las definiciones de los atributos asociados y las condiciones de su utilización. Este proceso requiere una gestión.
- **Identificar múltiples instancias de la misma entidad** representadas dentro y a través de las fuentes de datos; construir y mantener identificadores y referencias cruzadas para permitir la integración de la información.
- **Conciliar y consolidar los datos entre las fuentes** para proporcionar un registro maestro o la mejor versión de la verdad. Los registros consolidados proporcionan una visión fusionada de la información entre sistemas y tratan de resolver las incoherencias en la denominación de los atributos y el valor de los datos.

Gestión de datos maestros

- **Identificar las instancias que no coinciden o se fusionan de forma incorrecta** y garantizar que se resuelven y se asocian correctamente con los identificadores.
- **Provisión de acceso a datos confiables** a través de aplicaciones, ya sea a través de lecturas directas, servicios de datos, o por medio de la replicación de datos a almacenes de datos transaccionales, de almacenamiento o analíticos.
- **Imponer el uso de los valores de los datos** maestros dentro de la organización. Este proceso también requiere una gestión de gobierno y de cambio para asegurar una perspectiva empresarial compartida.

Reference and Master Data

Definition: Managing reconciled and integrated data through stewardship and semantic consistency in support of enterprise-wide needs to share its data assets.

Goals:

1. Ensuring the organization has complete, consistent, current, authoritative Master and Reference Data across organizational processes.
2. Enabling Master and Reference Data to be shared across enterprise functions and applications.
3. Lowering the cost and reducing the complexity of data usage and integration through standards, common data models, and integration patterns.

Business Drivers

Inputs:

- Candidate Data Stores and Values
- Cross Functional Requirements
- Industry Standards
- Metadata
- Purchased Data and/or Open Data and Code Sets
- Business Rules

Activities:

1. Define Drivers and Requirements (P)
2. Evaluate and Assess Data Sources (P)
3. Define Architectural Approach (D)
4. Model Data (D)
5. Define Stewardship and Maintenance Processes (C)
6. Establish Governance Policies (C)
7. Implement Data Sharing / Integration Services (D,O)
 1. Acquire Data Sources for Sharing
 2. Publish Reference and Master Data

Deliverables:

- Master and Reference Data Requirements
- Data Models and Integration Patterns
- Reliable Reference and Master Data
- Reusable Data Services
- Data Exception Reports

Suppliers:

- Subject Matter Experts
- Data Stewards
- Application Developers
- Data Providers
- Business Analysts
- Infrastructure Systems Analysts

Participants:

- Data Analysts
- Data Modelers
- Data Stewards
- Data Integrators
- Data Architects
- Data Quality Analysts

Consumers:

- Subject Matter Experts
- Data Integrators
- Application Users
- Application Developers
- Solution Architects

Technical Drivers

Techniques:

- Conditions-of-use agreements
- Business key cross references
- Processing Log analysis

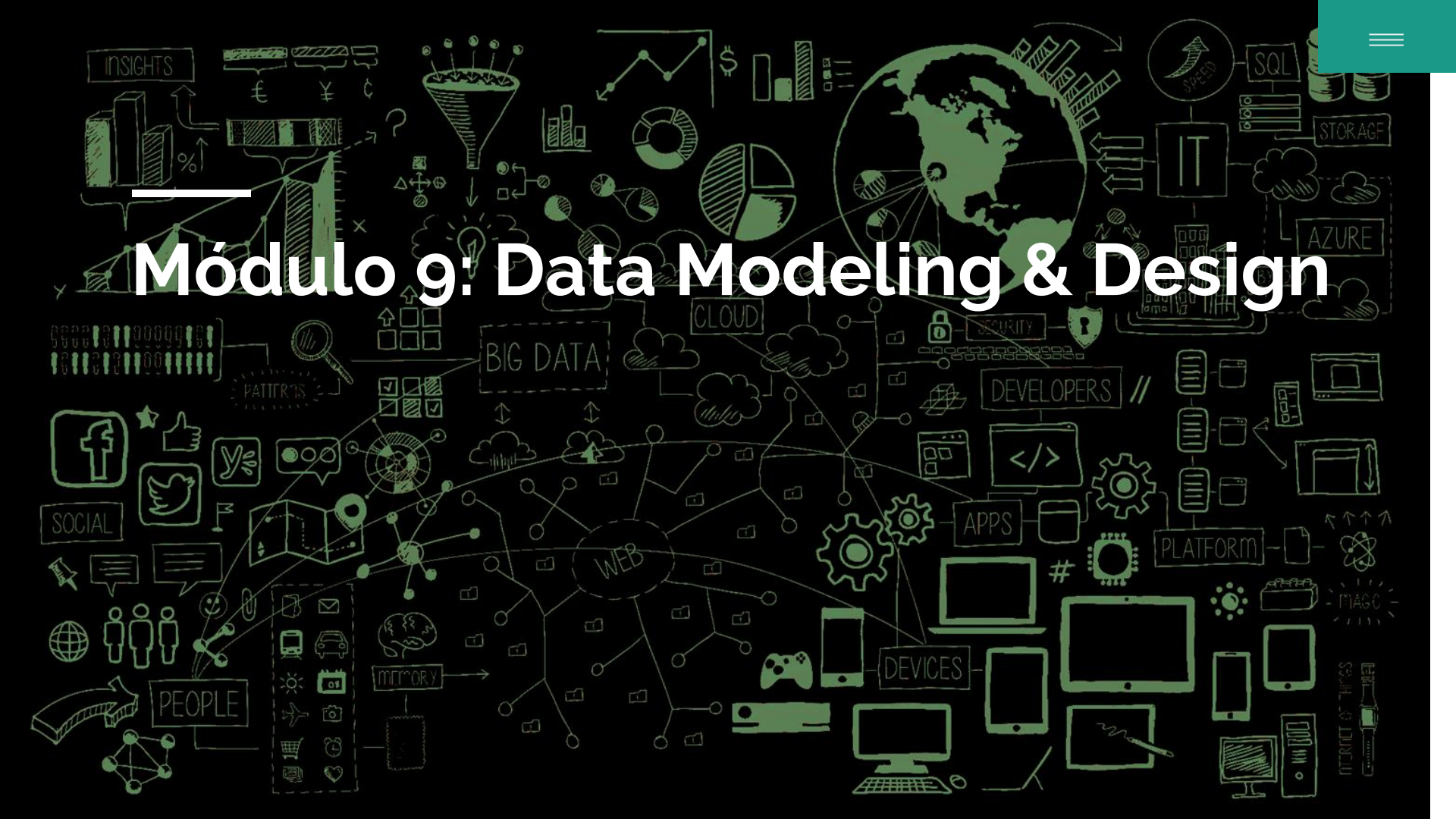
Tools:

- Data Integration Tools
- Data Remediation Tools
- Operational Data Stores
- Data Sharing Hubs
- Data Modeling Tools
- Metadata Repositories
- MDM Application Platforms

Metrics:

- Data Quality and Compliance
- Data Change Activity
- Data Ingestion and Consumption
- Data Sharing Availability
- Data Steward Coverage
- Data Sharing Volume and Usage

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



Módulo 9: Data Modeling & Design

Entidades

- Son el metadato básico, son un algo sobre lo que la organización recaba datos
- Puede considerarse como la respuesta a una pregunta fundamental: Quién, Qué, Cuándo, Dónde, Por Qué, Cómo o a una combinación de las mismas
- Suele representarse como un rectángulo



Definición de entidades

Al tratarse de metadatos fundamentales para el valor aportado al negocio por el modelo de datos, deben estar definidos de forma clara y con gran calidad:

- **Claridad:** La definición debe ser fácil de leer y comprender. Se recomienda utilizar frases simples y bien escritas sin acrónimos o términos ambiguos inexplicados. Evitar el uso de expresiones como “*a veces*” o “*normalmente*”
- **Precisión:** La definición debe ser una descripción precisa y correcta de la entidad. Deben ser revisadas por expertos de negocio para asegurar que sean precisas
- **Compleitud:** Todas las partes de la definición están detalladas. Por ejemplo, al definir código, se incluyen ejemplos de los valores del código. Al definir un identificador, se incluye en la definición el alcance de la unicidad

Atributos

Son una **propiedad que identifica, describe o mide a una entidad.**

Se suele representar como una lista bajo la entidad.

Producto
idproducto
codigobarras
categoria
marca
fechaalta
fechabaja
idproveedor
costecompra
precioventa

Relación

Es una asociación entre entidades

- Captura interacciones de alto nivel entre entidades conceptuales
- Identifica las interacciones detalladas entre entidades lógicas
- Indica las limitaciones entre entidades físicas



Cardinalidad de la relación

En la relación entre dos entidades la cardinalidad indica cuántas instancias de cada entidad participan a cada lado de la relación. Los valores que puede tener son 0, 1 o ‘muchos’

Cardinalidad	Uno	Muchos	Uno y sólo uno	Cero o uno	Uno o muchos	Cero o muchos
Representación	1	n	1..1	0..1	1..n	0..n
Símbolo	—+	—<	—⦿	—○+	—<	—○<

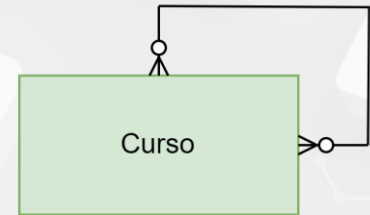
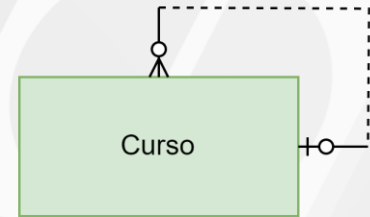


Valencia de la relación

Se llama valencia de una relación al **número de entidades** que forman parte de la misma

- **Relación unaria (recursiva)**

- Representa la relación de una entidad consigo misma
- Si la relación es algo a muchos representa una jerarquía
- Si la relación es de muchos a muchos representa una red o grafo
- En el primer ejemplo un curso puede tener 0 o n prerrequisitos, pero no puede a su vez ser requisito de otro
- En el segundo ocurre lo mismo pero el curso si que puede ser requisito



Valencias binarias y ternarias

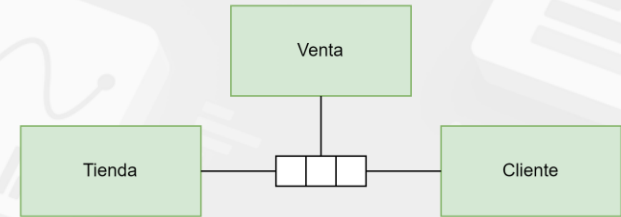
Valencia binaria

- Se da cuando una relación involucra 2 entidades
- Es, con mucho, la más común de las relaciones



Valencia ternaria

- Se da cuando una relación involucra 3 entidades
- Muy utilizada en el modelado por hechos



Claves

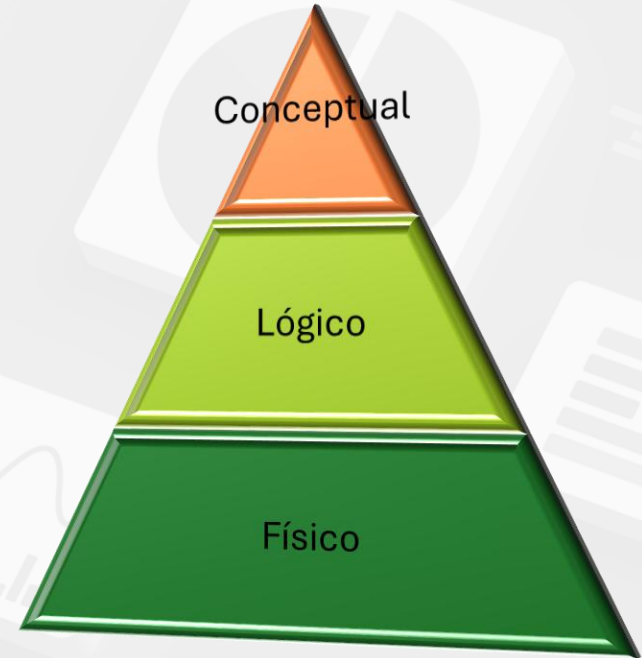
Una clave o identificador es un conjunto de uno o más atributos que definen de manera única una instancia de una entidad. Se pueden clasificar y tipificar de las siguientes formas:

- Por tipo de construcción
- Por tipo de función

Producto
idproducto
codigobarras
categoria
marca
fechaalta
fechabaja
idproveedor
costecompra
precioventa

Niveles de detalle del modelo de datos

- **Conceptual:** Representa requisitos de alto nivel como una colección de conceptos relacionados
- **Lógico:** Representación detallada de los requerimientos de datos, habitualmente en apoyo de un contexto específico
- **Físico:** Solución técnica detallada y adaptada al conjunto de hardware y software utilizado



Normalización

Proceso de aplicación de reglas para **crear estructuras de datos estables** manteniendo cada atributo en un lugar único eliminando redundancia e inconsistencias.

Cada nivel de normalización no tiene porqué incluir al anterior (aunque facilita su implementación) y **soluciona problemas adicionales**.

Se considera que los datos están **normalizados cuando están en 3NF**. Por encima de la 3NF las formas se consideran ejercicios académicos con una aplicación muy excepcional.

Normalización, formas normales

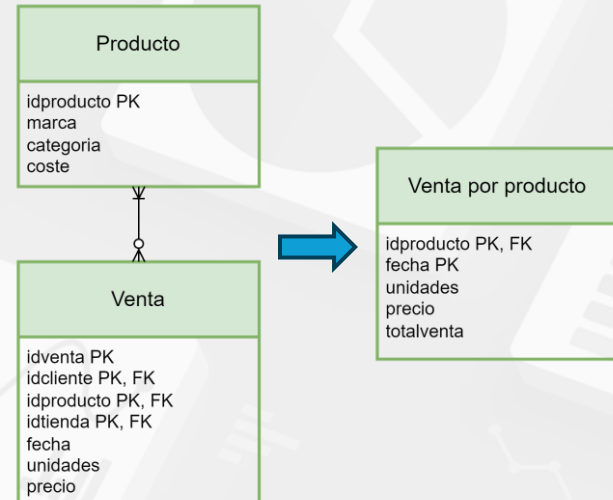
- **Primera forma normal (1NF):** Asegura que cada entidad tenga una clave primaria válida, y que cada atributo dependa de la clave primaria. Elimina grupos repetidos y asegura que cada atributo sea atómico (sin multi valor). 1NF incluye la resolución de relaciones de muchos a muchos con una entidad adicional a menudo llamada entidad asociativa o puente.
- **Segunda forma normal (2NF):** Asegura que cada entidad tiene una clave primaria mínima y que cada atributo depende de la clave primaria completa.
- **Tercera forma normal (3NF):** Asegura que cada entidad no tiene claves primarias ocultas y que cada atributo no depende de ningún atributo fuera de la clave. “La clave, la clave completa y nada más que la clave”.
- **Forma normal de Boyce Codd (BCNF):** Resuelve la superposición de claves candidatas compuestas. Clave superpuesta significa que hay reglas de negocio ocultas entre las claves.
- **Cuarta forma normal (4NF):** Resuelve todas las relaciones de muchos a muchos a muchos (y más allá) en pares hasta que no pueden ser desglosadas en bloques más pequeños.
- **Quinta forma normal (5NF):** Resuelve las dependencias entre entidades en pares básicos, y todas las dependencias unidas utilizan partes de las claves primarias.

Desnormalización

Transformación deliberada creando tablas con estructuras de datos redundantes o duplicados. Se ponen atributos en varios lugares de forma intencionada. Se hace para mejorar el rendimiento:

- Combinar datos de varias tablas para evitar joins
- Crear copias de datos previamente filtrados
- Precálculo y almacenamiento de datos

Se introduce un riesgo de error por duplicados y hace más complejo el mantenimiento para mantener los datos actualizados





Plan de modelado de datos

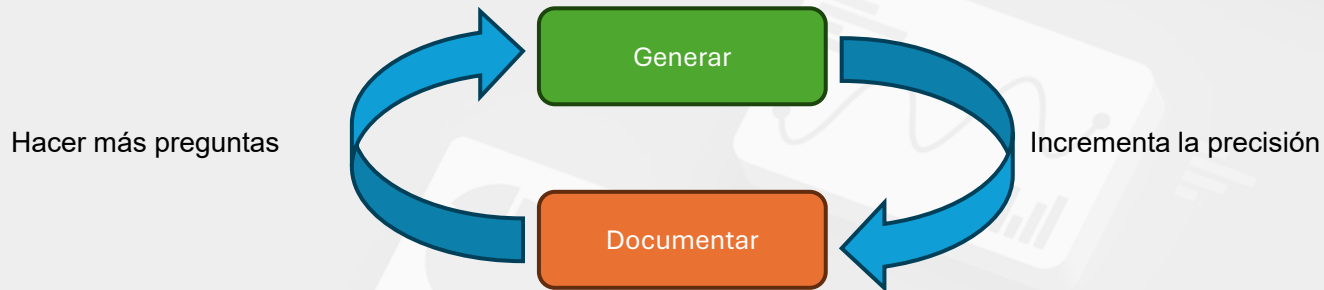
Debe contener tareas como **evaluar requerimientos organizacionales**, creando estándares y definiendo el almacenamiento.

Puede incluir los entregables siguientes:

- **Diagrama:** Forma visual que recoge los requerimientos de forma precisa. Se define por nivel de detalle, tipo de esquema y notación
- **Definiciones:** Descripción de las entidades, atributos y relaciones
- **Problemas y cuestiones pendientes:** Listado de cuestiones pendientes de resolución, habitualmente por falta de contexto de negocio. Se resolverán en iteraciones posteriores y durante el propio proceso de modelado
- **Linaje:** Mapeos entre origen y destino. Relación entre componentes de los modelos conceptual, lógico y físico

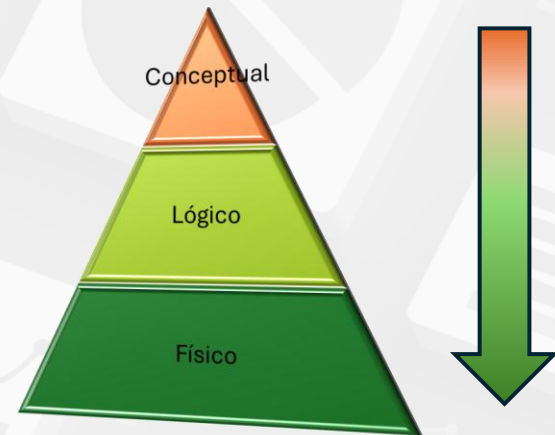
Construir el modelo de datos

- La construcción de un modelo es un proceso iterativo
- Se prepara una versión y se coteja con el negocio
- Se revisa el modelo, se generan nuevas preguntas y el ciclo comienza de nuevo



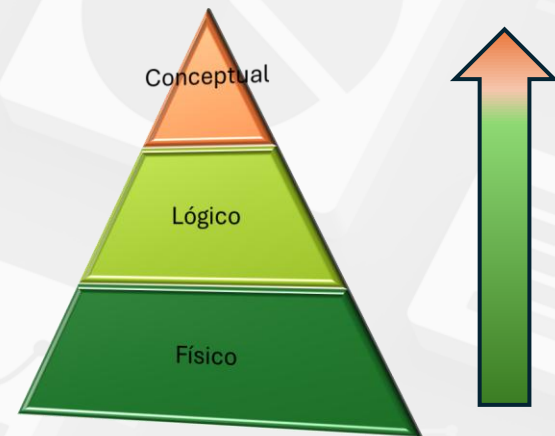
Construir el modelo, ingeniería directa

- Es el proceso de documentar partiendo de los requerimientos
- Se comienza con el CDM
- El LDM se crea para documentar la solución de negocio
- Finalmente se trabaja con el PDM para definir la parte técnica



Construir el modelo, ingeniería inversa

- Es el proceso de documentar una base de datos existente
- Se comienza con el PDM
- El LDM se crea o actualiza partiendo del PDM
- Finalmente se trabaja con el CDM tomando el LDM como base



Patrones y modelos de datos de industria

- **Patrones de modelo de datos:** Estructuras de modelado y patrones de integración reutilizables aplicables para solventar ciertas situaciones. Por ejemplo, permiten resolver relaciones de muchos a muchos o ayudan a construir jerarquías auto referenciadas.
- **Modelos de datos de industria:** Son modelos de datos preconstruidos para todo un sector, como salud o banca. Son de amplio alcance, muy detallados y se deben adaptar y ajustar a la realidad de cada organización. Pueden adquirirse a través de proveedores o grupos industriales. Varios ejemplos: ARTS (comercio minorista), SID (telecomunicaciones) o ACORD (seguros)

Data Modeling and Design

Definition: Data modeling is the process of discovering, analyzing, and scoping data requirements, and then representing and communicating these data requirements in a precise form called the data model. This process is iterative and involves conceptual, logical, and physical models.

Goals:

1. To confirm and document an understanding of different perspectives, which leads to applications that more closely align with current and future business requirements.
2. To understand how data fits together and creates a foundation to successfully complete broad-scoped initiatives such as master data management and data governance.
3. To lower support costs and reduce the costs of building new applications thanks to increased reusability opportunities.

Business Drivers

Inputs:

- Enterprise Data Model
- Existing Data Models and Databases
- Data Standards
- Data Sets
- Data Requirements
- Data Architecture
- Enterprise Taxonomy

Activities:

1. **Plan for Data Modeling (P)**
2. **Build the Data Model (D)**
 1. Conceptual Data Modeling
 2. Logical Data Modeling
 3. Physical Data Modeling
3. **Review the Data Models (C)**
4. **Maintain the Data Models (O)**

Deliverables:

- Diagrams (different levels, schemes, and notations)
- Definitions
- Issues and outstanding questions
- Lineage

Suppliers:

- Business Professionals
- Business Analysts
- Data Architects
- Database Administrators and Developers
- Subject Matter Experts
- Data Stewards
- Metadata Administrators

Participants:

- Business Analysts
- Data Analysts
- Data Modelers
- Data Consumers

Consumers:

- Business Analysts
- Data Modelers
- Database Administrators and Developers
- Software Developers
- Data Stewards
- Data Quality Analysts
- Data Consumers

Technical Drivers

Techniques:

- Naming conventions
- Forward/reverse engineering
- Approach determination

Tools:

- Data modeling tools
- Lineage tools
- Data profiling tools
- Metadata repositories
- Data model patterns
- Industry data models

Metrics:

- Data model scorecard

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



Módulo 10: Data Quality

Ciclo de vida del Dato



Ciclo de vida del Dato

El dato normalizado y de calidad está a la base de cualquier uso del mismo. Sin datos de cualidad no hay información fiable y consistente. Estas son las 6 fases del ciclo de vida del Dato

- **Creación:** Implica la generación de nuevo contenido digital, o la alteración, actualización o modificación de un contenido existente.
- **Almacenamiento:** el acto de ubicar los datos digitales en algún tipo de repositorio para su custodia y normalmente ocurre de forma prácticamente simultánea a su creación.
- **Uso.** Los datos son visualizados, procesados, o utilizados de otro modo en algún tipo de actividad, no incluyendo su modificación.
- **Condivisión.** La información se hace accesible a otros, tales como otros usuarios, clientes, proveedores, etc.

IMPORTANTE EXAMEN

Archivado. Los datos dejan de ser usados activamente y pasan a conservarse en un almacenamiento de largo plazo.

Destrucción. Los datos son destruidos de forma permanente usando medios físicos o lógicos. Es en esta fase donde se cierra el ciclo de vida del dato o contenido digital, puesto que tras esta fase ya no estará accesible.

Los datos tienen un fin y un final. El fin es ser un activo de la organización el final es cuando ya no son necesario e incluso su custodia puede ser perjudicial o lesiva. Contrariamente a lo que se piensa no tiene sentido conservar todos los datos a futuro. Sobre todo porque la compliance nos lo impide. Aunque el coste de almacenamiento ya no es crítico, hay que librarse de los datos cuando estos ya no son un activo.

Data Quality: Métricas de calidad de datos efectivas

Gran parte del trabajo de un equipo de calidad de datos se centrará en la medición y la presentación de informes sobre la calidad. Entre las categorías de alto nivel de la medición de la calidad de los datos se incluyen:

- **Retorno de la inversión:** Declaraciones sobre el costo de los esfuerzos de mejora frente a los beneficios de la mejora de la calidad de los datos
- **Niveles de calidad:** Mediciones del número y porcentaje de errores o violaciones de los requisitos dentro de un conjunto de datos o entre conjuntos de datos
- **Tendencias de la calidad de los datos:** Mejora de la calidad a lo largo del tiempo (es decir, una tendencia) con respecto a los umbrales y objetivos, o incidentes de calidad por período

Métricas de gestión de problemas de datos:

- Recuento de problemas por dimensiones de la calidad de los datos
- Cuestiones por función comercial y sus estados (resuelto, pendiente, escalado)
- Problema por prioridad y gravedad Tiempo para resolver los problemas

Dimensiones de Calidad de Datos

DAMA Reino Unido elaboró un White Paper que describe seis dimensiones fundamentales de la calidad de los datos:

- **Compleitud:** La proporción de datos almacenados contra el potencial del 100%.
- **Unicidad:** Ninguna instancia (cosa) de la entidad se registrará más de una vez basándose en cómo se identifica esa cosa.
- **Puntualidad:** El grado en que los datos representan la realidad desde el momento requerido.
- **Validez:** Los datos son válidos si se ajustan a la sintaxis (formato, tipo, rango) de su definición.
- **Precisión:** El grado en que los datos describen correctamente el objeto o evento del "mundo real" que se describe.

- **Consistencia:** La ausencia de diferencia, cuando se comparan dos o más representaciones de una cosa con una definición.
- **Usabilidad:** ¿Los datos son compresibles, simples, relevantes, accesibles, mantenibles y con el nivel de precisión adecuado?
- **Cuestiones de tiempo (más allá de la propia puntualidad):** ¿Es estable y a la vez responde a las solicitudes de cambio legítimas?
- **Flexibilidad:** ¿Son los datos comparables y compatibles con otros datos? ¿Tiene agrupaciones y clasificaciones útiles? ¿Pueden ser reutilizados? ¿Es fácil de manipular?
- **Confianza:** ¿Se han establecido procesos de gobernanza, protección y seguridad de los datos? ¿Cuál es la reputación de los datos, y es verificada o comprobable?
- **Valor:** ¿Existe un buen caso de costo/beneficio para los datos? ¿Se está utilizando de manera óptima? ¿Pone en peligro la seguridad o la privacidad de las personas, o las responsabilidades legales de la empresa? ¿Apoya o contradice el la imagen corporativa o el mensaje corporativo?



Iso 8000

ISO 8000 es el estándar global para la Calidad de los Datos y los Datos Maestros de la Empresa.

- Describe las características y define los requisitos para el intercambio estándar de Datos Maestros entre los socios comerciales.
- Establece el concepto de Portabilidad como requisito para los Datos Maestros de la Empresa
- Los verdaderos Datos Maestros de la Empresa son únicos para cada organización.

IMPORTANTE EXAMEN

Data Quality Management

Definition: The planning, implementation, and control of activities that apply quality management techniques to data, in order to assure it is fit for consumption and meets the needs of data consumers.

Goals:

1. Develop a governed approach to make data fit for purpose based on data consumers' requirements.
2. Define standards, requirements, and specifications for data quality controls as part of the data lifecycle.
3. Define and implement processes to measure, monitor, and report on data quality levels.
4. Identify and advocate for opportunities to improve the quality of data, through process and system improvements.

Business Drivers

Inputs:

- Data Policies and Standards
- Data Quality Expectations
- Business Requirements
- Business Rules
- Data Requirements
- Business Metadata
- Technical Metadata
- Data Sources and Data Stores
- Data Lineage

Activities:

1. **Define High Quality Data (P)**
2. **Define a Data Quality Strategy (P)**
3. **Define Scope of Initial Assessment (P)**
 1. Identify Critical Data
 2. Identify Existing Rules and Patterns
4. **Perform Initial Data Quality Assessment (P)**
 1. Identify and prioritize issues
 2. Perform root cause analysis of issues
5. **Identify & Prioritize Improvements**
 1. Prioritize Actions based on Business Impact
 2. Develop Preventative and Corrective Actions
 3. Confirm Planned Actions
6. **Develop and Deploy Data Quality Operations (D)**
 1. Develop Data Quality Operational Procedures
 2. Correct Data Quality Defects
 3. Measure and Monitor Data Quality
 4. Report on Data Quality levels and findings.

Deliverables:

- Data Quality Strategy & framework
- Data Quality Program organization
- Analyses from Data Profiling
- Recommendations based on root cause analysis of issues
- DQM Procedures
- Data Quality Reports
- Data Quality Governance Reports
- Data Quality Service Level Agreements
- DQ Policies and Guidelines

Suppliers:

- Business Management
- Subject Matter Experts
- Data Architects
- Data Modelers
- System Specialists
- Data Stewards
- Business Process Analysts

Participants:

- CDO
- Data Quality Analysts
- Data Stewards
- Data Owners
- Data Analysts
- Database Administrators
- Data Professionals
- DQ Managers
- IT Operations
- Data Integration Architects
- Compliance Team

Consumers:

- Business Data Consumers
- Data Stewards
- Data Professionals
- IT Professionals
- Knowledge Workers
- Data Governance Bodies
- Partner Organizations
- Centers of Excellence

Technical Drivers

Techniques:

- Spot-Checking using Multiple Subsets
- Tags and Notes to Mark Data Issues
- Root Cause Analysis
- Statistical Process Control

Tools:

- Profiling engines, query tools
- Data Quality Rule Templates
- Quality Check and Audit Code Modules

Metrics:

- Governance and Conformance Metrics
- Data Quality Measurement Results
- Improvement trends
- Issue Management Metrics

Data Security

La seguridad de los datos incluye la planificación, el desarrollo y la ejecución de políticas y procedimientos de seguridad para proporcionar una autenticación, autorización, acceso y auditoría adecuados de los datos y los activos de información.

Los aspectos clave de Data Security son:

- **Los Stakeholders (Grupo de Interés):** Las organizaciones deben reconocer las necesidades de privacidad y confidencialidad de sus interesados, incluidos los clientes, pacientes, estudiantes, ciudadanos, proveedores o socios comerciales. Todos los integrantes de una organización deben ser fideicomisarios responsables de los datos sobre las partes interesadas.
- **Regulaciones gubernamentales:** Se han establecido reglamentos gubernamentales para proteger los intereses de algunos interesados. Los reglamentos tienen diferentes objetivos. Algunos restringen el acceso a la información, mientras que otros garantizan la apertura, la transparencia y la responsabilidad. (GDPR, Basilea, Solvency)

Data Security

- **Assets que son parte de la ventaja competitiva de la empresa:** Cada organización tiene datos de propiedad exclusiva que debe proteger. Los datos de una organización permiten conocer a sus clientes y, cuando se aprovechan eficazmente, pueden proporcionar una ventaja competitiva. Si los datos confidenciales son robados o violados, una organización puede perder la ventaja competitiva.
- **Necesidades de acceso legítimo:** Al proteger los datos, las organizaciones también deben permitir el acceso legítimo. Los procesos empresariales requieren que las personas que desempeñan determinadas funciones puedan acceder a los datos, utilizarlos y mantenerlos.
- **Obligaciones contractuales:** Los acuerdos contractuales y de no divulgación también influyen en los requisitos de seguridad de los datos. Por ejemplo, la norma PCI, un acuerdo entre las empresas de tarjetas de crédito y las empresas comerciales individuales, exige que se protejan determinados tipos de datos de maneras definidas (por ejemplo, la codificación obligatoria de las contraseñas de los clientes).

Distinciones en materia de seguridad

1

Vulnerabilidad

Una vulnerabilidad es una debilidad o defecto en un sistema que permite que sea atacado y comprometido con éxito - esencialmente un agujero en las defensas de una organización.

2

Amenazas

Una amenaza es una acción ofensiva potencial que podría ser tomada contra una organización. Las amenazas pueden ser internas o externas.

3

Riesgo

El término riesgo se refiere tanto a la posibilidad de pérdida y también a la condición que plantea la pérdida potencial.

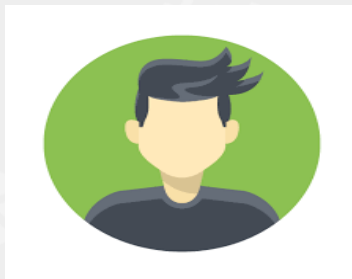
¿De quién es la responsabilidad?

Chief Information Security Officer (CISO)



- Asegurar la información desde el punto de vista de seguridad y protección de los datos
- Definir y mantener procesos para garantizar el correcto uso de los datos
- Asegurar la confidencialidad de los mismos
- Determinar las medidas de protección de datos
- Definir criterios de confidencialidad a nivel de metadatos

Chief Data Officer (CDO)



- Gestionar el movimiento y consolidación de datos
- Velar por la aplicación de los requisitos y criterios de seguridad y protección de los datos
- Definir y mantener atributos para la clasificación de confidencialidad
- Conocer la definición y aplicación de los requerimientos regulatorios relativos a los datos
- Implantar las políticas, regulaciones y normativas referentes a los datos

Data Protection Officer (DPO)



- Definir y mantener procesos y estándares para garantizar la clasificación de los datos sensibles y asegurar la privacidad de los mismos
- Definir los criterios de privacidad de los metadatos asociados
- Coordinar con la data office la implantación de las políticas, regulaciones y normativas referentes a los datos

El enfoque de DAMA DMBOK

La seguridad de la información comienza por clasificar los datos de una organización a fin de identificar qué datos requieren protección. El proceso general incluye los siguientes pasos:

- **Identificar y clasificar los activos de datos sensibles:** Dependiendo de la industria y la organización, puede haber pocos o muchos activos y una gama de datos sensibles (incluyendo la identificación personal, médica, financiera y más).
- **Localizar los datos sensibles en toda la empresa:** Los requisitos de seguridad pueden ser diferentes, dependiendo de dónde se almacenen los datos. Una cantidad significativa de datos sensibles en una sola ubicación supone un alto riesgo debido al daño posible de una sola brecha.
- **Determinar cómo se debe proteger cada activo:** Las medidas necesarias para garantizar la seguridad pueden variar entre los bienes, según el contenido de los datos y el tipo de tecnología.
- **Identificar la forma en que esta información interactúa con los procesos comerciales:** Es necesario analizar los procesos comerciales para determinar qué acceso se permite y en qué condiciones.

Las 4 As de la seguridad de datos

- 1. Autenticación:** Validar el acceso de los usuarios. Cuando un usuario intenta acceder a un sistema, el sistema necesita verificar que la persona es quien dice ser.
 - **Ejemplo:** Uso de contraseñas, biometría (huellas dactilares, reconocimiento facial), tarjetas inteligentes, autenticación de dos factores (2FA).
- 2. Autorización:** Proceso que determina qué recursos y acciones están permitidos para un usuario autenticado.
 - **Ejemplo:** Un usuario con privilegios de administrador puede acceder a todos los archivos del sistema, mientras que un usuario estándar solo puede acceder a sus propios archivos.
- 3. Auditoría:** Proceso de registrar y revisar actividades y eventos para asegurar que se cumplen las políticas de seguridad y para identificar actividades sospechosas o no autorizadas.
 - **Ejemplo:** Registros de acceso, registros de cambios en los sistemas, análisis de logs para detectar intentos de intrusión.
- 4. Administración:** Gestión y mantenimiento continuo de las políticas, procedimientos y herramientas de seguridad para garantizar que el sistema de seguridad funcione correctamente.
 - **Ejemplo:** Configuración de controles de seguridad, actualización de software de seguridad, capacitación de usuarios sobre buenas prácticas de seguridad, gestión de usuarios y permisos.

CRUDE Matrix, Metadata Tracking, Data Masking - Pseudo Anonymization

El framework de DAMA DmBok 2 enfoca el riesgo de datos con esta fórmula:

- Vulnerabilidad
- Amenaza
- Riesgo
 - Critical Risk Data (CRD)
 - High Risk Data (HRD)
 - Moderate Risk Data (MDR)

Crude Matrix

Esta matriz se utiliza en Data Security para vigilar los procesos de relación **(CRUD– Create,Read, Update, Delete)** ayuda a mapear las necesidades de acceso a los datos y orientan la definición de los grupos de funciones de seguridad de los datos, los parámetros y los permisos. Algunas versiones añaden una E para “Ejecutar” y se transforma en “CRUDE.”

CRUD MATRIX							
Calling Item	Item Type	COUNTER	CUSTOMER	EMPLOYEE	PRODUCT	SALES_ORDER_ITEMS	SALES_ORDER
ALL_SALES_ORDER_ITEMS_DETAILS	View			R	R	R	R
ALL_PRODUCTS	View				R		
ALL_ORDERS_BY_EMPLOYEE	View						R
EMPLOYEE_COUNT	Trigger			R			
TRG_ORDER	Trigger				R		
CUSTOMERS.InsertCustomer	Procedure		C				
CUSTOMERS.UpdateCustomerName	Procedure		U				
CUSTOMERS.DeleteCustomerById	Procedure		D				
EMPLOYEES.OrdersByEmployee	Procedure		R	R		R	R
PRODUCTS.ProductsByCustomer	Procedure		R		R	R	R
PRODUCTSBYCUSTOMER	Procedure		R		R	R	R
pfc_updateprep	Event	RU					
itemchanged	Event				R		
demopfc.pbl.d_tab_customer	Datawindow Object		RU				
demopfc.pbl.d_tab_sales_order	Datawindow Object		R	R	R	R	R
demopfc.pbl.d_ff_sales_order	Datawindow Object		R	R			RU
demopfc.pbl.d_tab_employee	Datawindow Object			RU			

Table or Column Accessed

Components Accessing the table/column

Type of Access (Create, Read, Update, Delete)

Data Security

Definition: The planning, development, and execution of security policies and procedures to provide proper authentication, authorization, access, and auditing of data and information assets within cultural and regulatory considerations.

Goals:

1. Enable appropriate, and prevent inappropriate, access to enterprise data assets.
2. Understand and comply with all relevant regulations and policies for privacy, protection, and confidentiality.
3. Ensure that the privacy and confidentiality needs of all stakeholders are enforced and audited.

Business Drivers

Inputs:

- Business goals and strategy
- Business rules and processes
- Regulatory requirements
- Enterprise Architecture standards
- Enterprise Data Model
- Cybersecurity updates

Activities:

1. Identify Relevant Data Security Requirements (P)
2. Define Data Security Policy (C)
3. Define Data Security Standards (D)
4. Assess Current Security Risks (P)
5. Implement and Controls and Procedures (O)

Deliverables:

- Data security architecture
- Data security policies
- Data privacy and confidentiality standards
- Data security access controls
- Regulatory compliant data access views
- Documented security classifications
- Authentication and user access history
- Data security audit reports

Suppliers:

- IT Steering Committee
- Enterprise Architects
- Government
- Regulatory Bodies

Participants:

- Data Stewards
- Information Security Team
- Internal Auditors
- Process Analysts

Consumers:

- Business Users
- Regulatory Auditors

Technical Drivers

Techniques:

- CRUD Matrix Usage
- Immediate Security Patch
- Data Security Attributes in Metadata
- Security Needs in Projects
- Efficient Search for Encrypted Data
- Document Sanitization

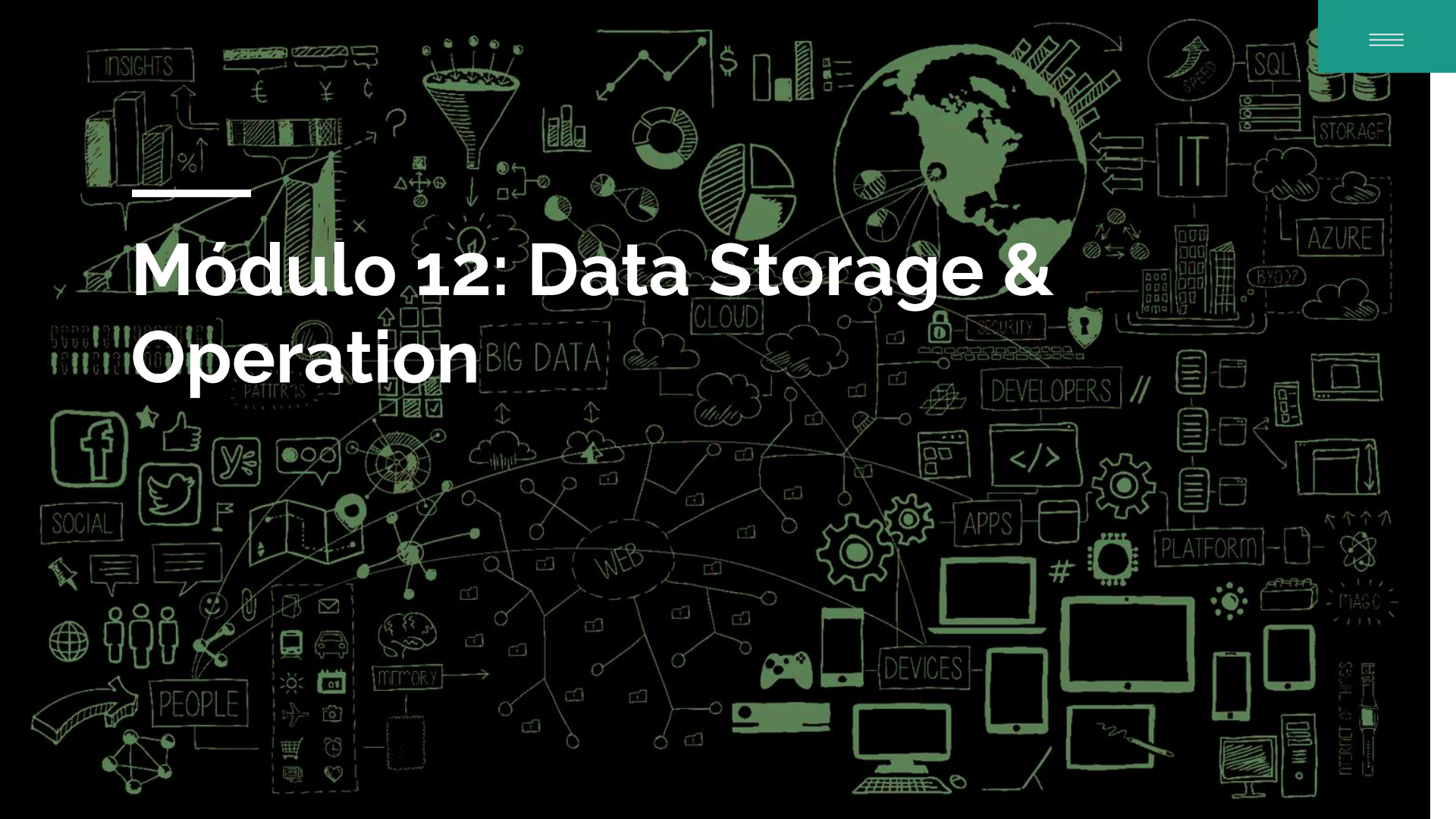
Tools:

- Anti-virus and Security Software
- Webpage Security
- Identity Management Technology
- Intrusion Detection and Prevention Software
- Firewalls
- Metadata tracking
- Data Masking / Encryption

Metrics:

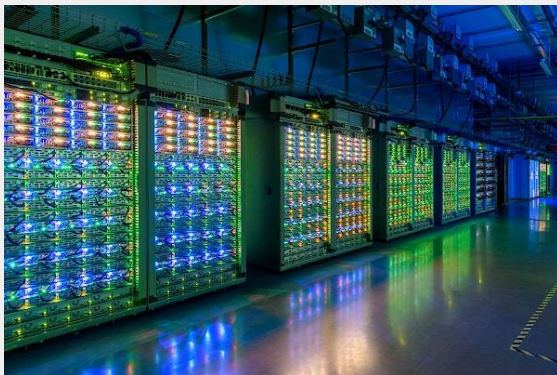
- Security Implementation Metrics
- Security Awareness Metrics
- Data Protection Metrics
- Security Incident Metrics
- Confidential Data Proliferation Rate

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



Módulo 12: Data Storage & Operation

Almacenamiento de datos y operaciones



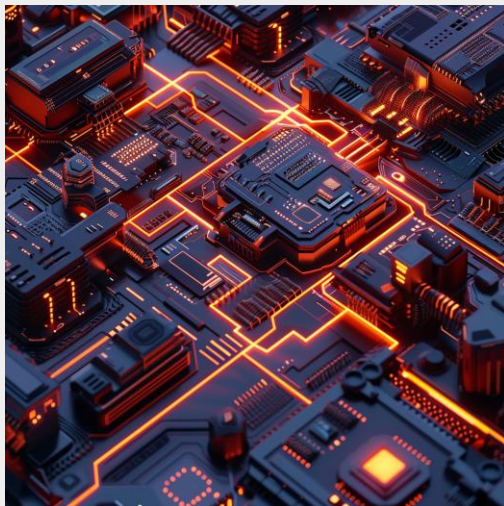
- Es el diseño, la implantación y soporte de los datos almacenados con el objetivo de maximizar su valor a lo largo de todo su ciclo de vida. Incluye:
 - El soporte de BB.DD
 - El soporte tecnológico de las BB.DD

Metas



- Gestionar la **disponibilidad** de los datos a lo largo de su ciclo de vida
- Garantizar la **integridad** de los activos de datos
- Gestionar el **rendimiento** de las transacciones de datos

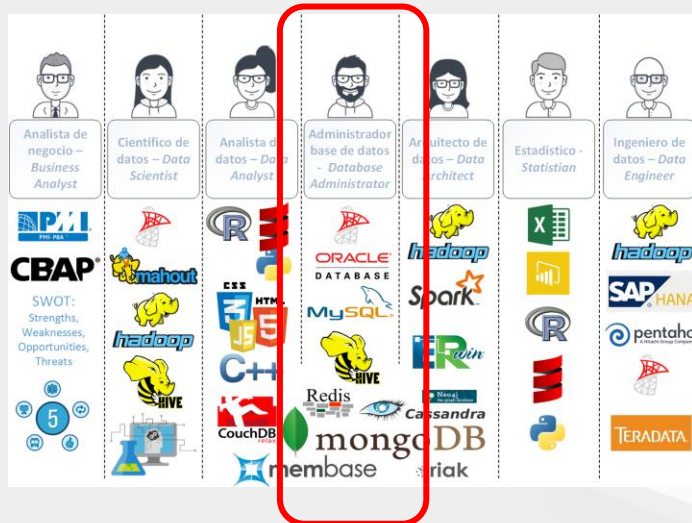
Términos y conceptos de BB.DD



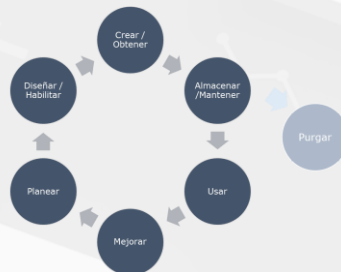
Base de datos: Cualquier colección de datos almacenados, independientemente de su estructura o contenido. Algunas grandes bases de datos se refieren a **instancias y a esquemas**.

- **Instancia:** Una ejecución del software de la base de datos que controla el acceso a una determinada área de almacenamiento. **Una organización suele tener varias instancias que se ejecutan simultáneamente**, utilizando diferentes áreas de almacenamiento. **Cada instancia es independiente de las demás.**
- **Esquema:** Un subconjunto de objetos de una base de datos contenidos en la base de datos o en una instancia. **Los esquemas se utilizan para organizar los objetos en partes más manejables.** Un esquema tiene un propietario y una lista de acceso particular para el contenido del esquema. El esquema también puede utilizarse para referirse a una colección de estructuras de bases de datos con algo en común.

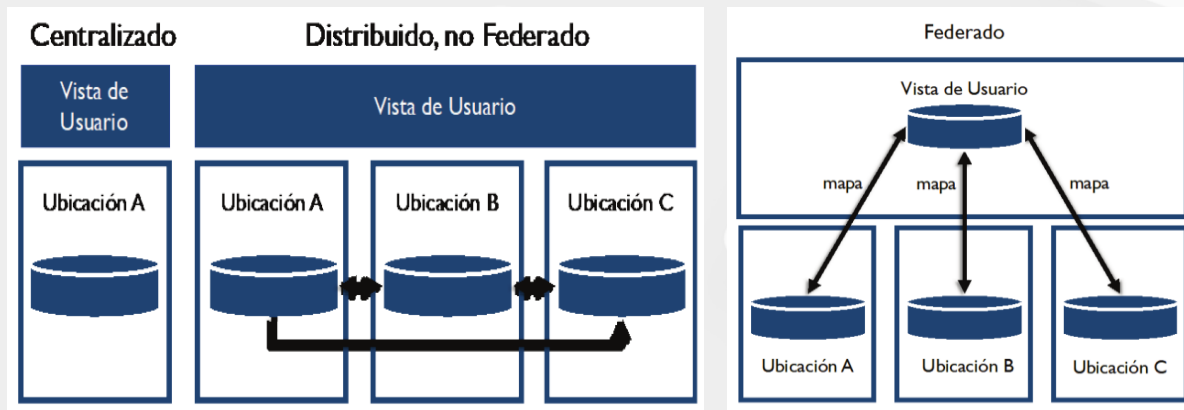
Gestión del ciclo de vida de los datos

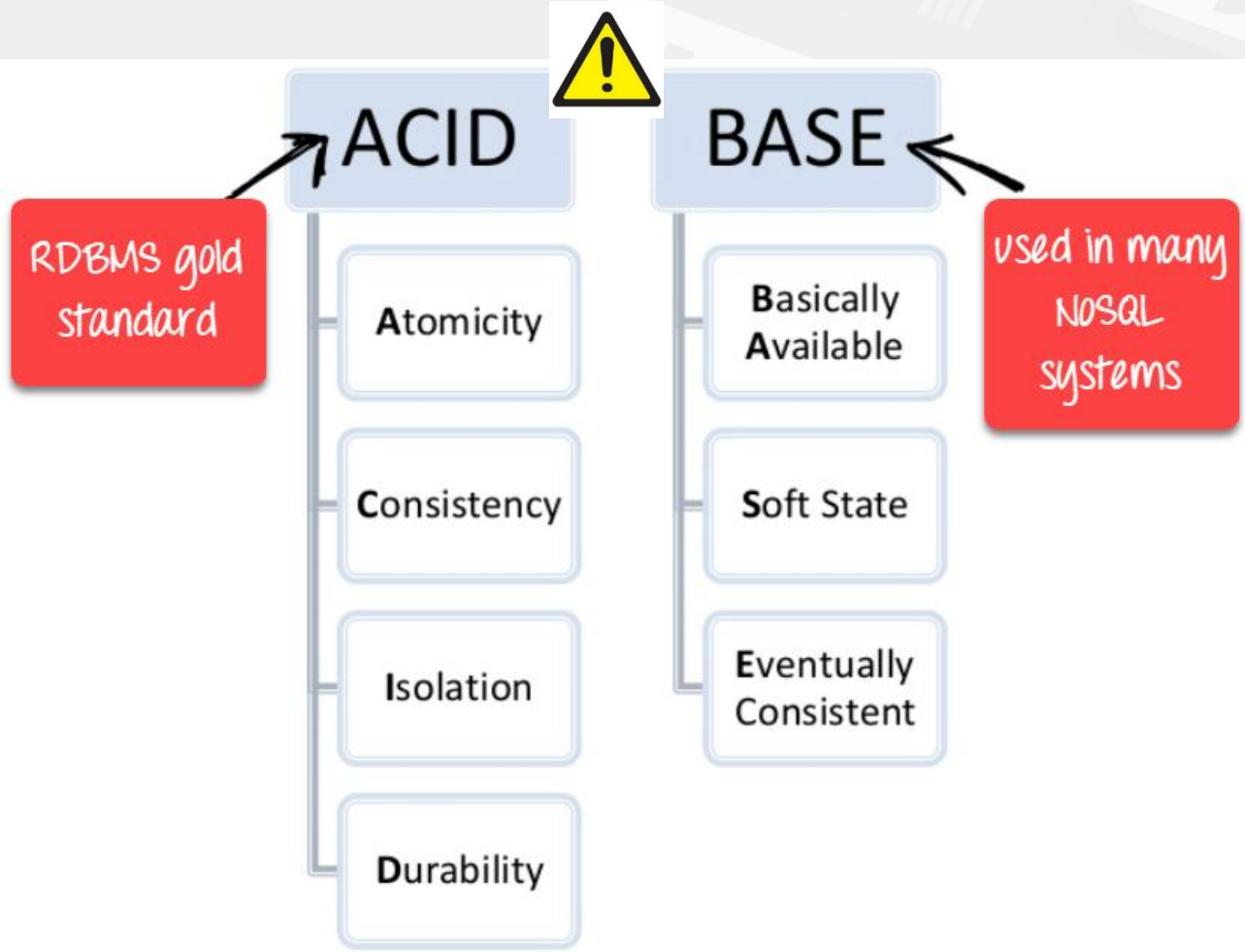


- Los DBAs **mantienen y aseguran la precisión y consistencia de los datos durante todo su ciclo de vida** a través del diseño, implementación y uso de cualquier sistema que almacene, procese o recupere datos.
- El DBA es el custodio de todos los cambios en las BB.DD.
- La gestión del ciclo de vida de los datos incluye la implemetación de políticas y procedimientos para la adquisición, migración, retención, expiración y disposición de los datos:

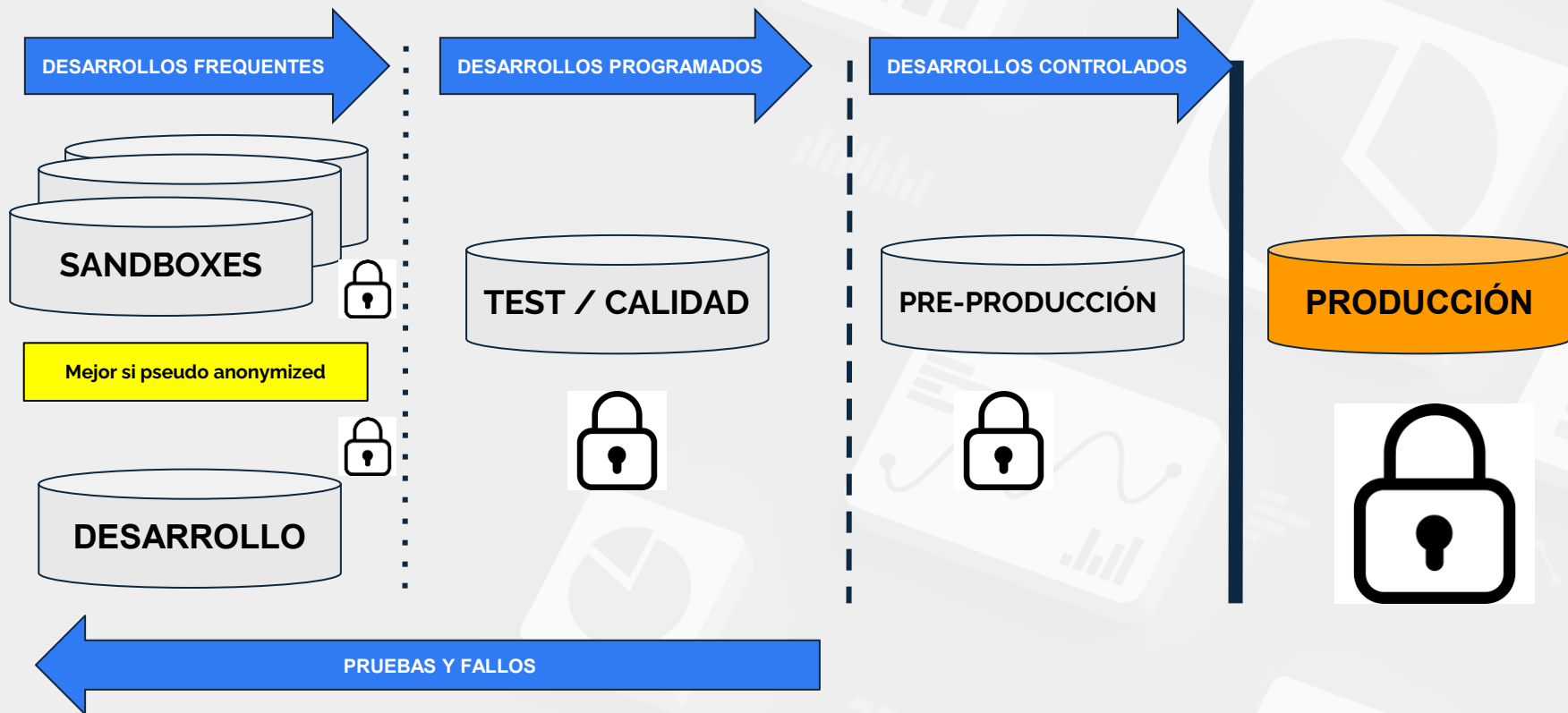


Tipos de arquitecturas de BB.DD





Entornos de base de datos





#GCPsketchnotes



@PVERGADIA



THECLOUDGIRL.DEV

07.10.2021

Which Database should I use?



RELATIONAL



Cloud SQL

Managed MySQL, PostgreSQL, SQL Server



Cloud Spanner

Cloud-native with large scale, consistency, 99.999% availability



Bare Metal

Lift and shift Oracle workloads to Google Cloud

Good For:

General purpose SQL DB

RDBMS+ scale, HA, HTAP

RDBMS+ scale, HA, HTAP

Use Case:



Web frameworks



Gaming



Legacy applications



ERP



Global financial ledger



Data center retirement



CRM



Ecommerce and web



Supply chain/inventory management



SaaS application

NON-RELATIONAL (NO SQL)

DOCUMENT



Firestore

Cloud Native, serverless, NoSQL document database, backend-as-a-service, global strong consistency, 99.999% SLA

KEY VALUE



Cloud Bigtable

Cloud-native NoSQL wide-column store for large scale, low-latency workloads

Good For:

Large scale, complex hierarchical data

Heavy read + write, events

Use Case:



Mobile/web/IoT applications



Personalization



Real-time sync



Adtech



Offline sync



Recommendation engines



Personalized apps



Fraud detection

IN MEMORY



Memory Store

Fully managed Redis and Memcached for sub-millisecond data access

Good For:

In-memory and Key-value store

Use Case:



Caching



Session store



Gaming



Personalization



Leaderboard



Adtech



Social chat or news feed

Cortesía de Google

Data Storage and Operations Governance

Métricas de DS&O

- **Conteos de BB.DD.**
- **Estadísticas sobre el número de transacciones**
- **Métricas de capacidad:**
 - Capacidad de almacenamiento ocupado
 - Número de objetos en la BB.DD.
- **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

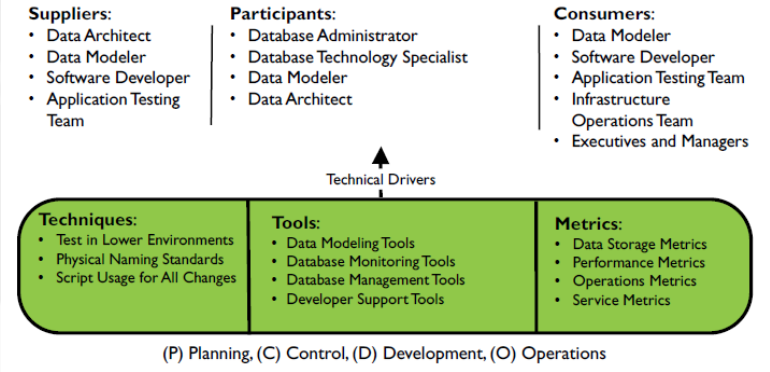
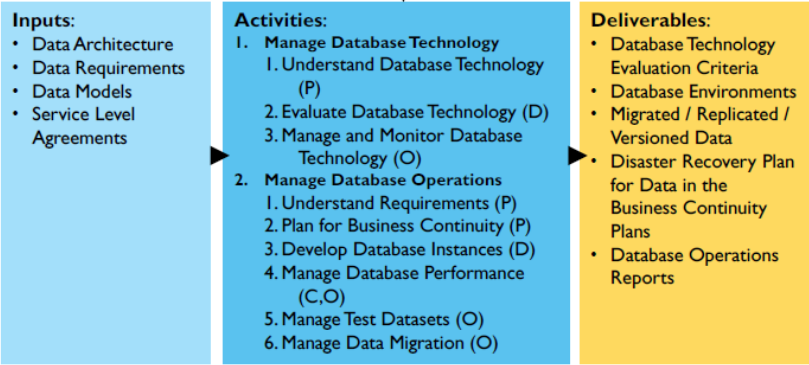


Data Storage and Operations

Definition:The design, implementation, and support of stored data to maximize its value.

- Goals:**
- 1. Manage availability of data throughout the data lifecycle.
 - 2. Ensure the integrity of data assets.
 - 3. Manage performance of data transactions.

Business Drivers





Módulo 13: DW & Business Intelligence

¿Qué es Business Intelligence (BI)?



“Tiene dos significados. En primer lugar, se refiere a un tipo de análisis de datos destinado a comprender las actividades y oportunidades de la organización. Los resultados de este análisis se utilizan para impulsar el éxito de la organización. [...] En segundo lugar, BI se refiere a un conjunto de tecnologías que soportan este tipo de análisis de datos.”

Gartner

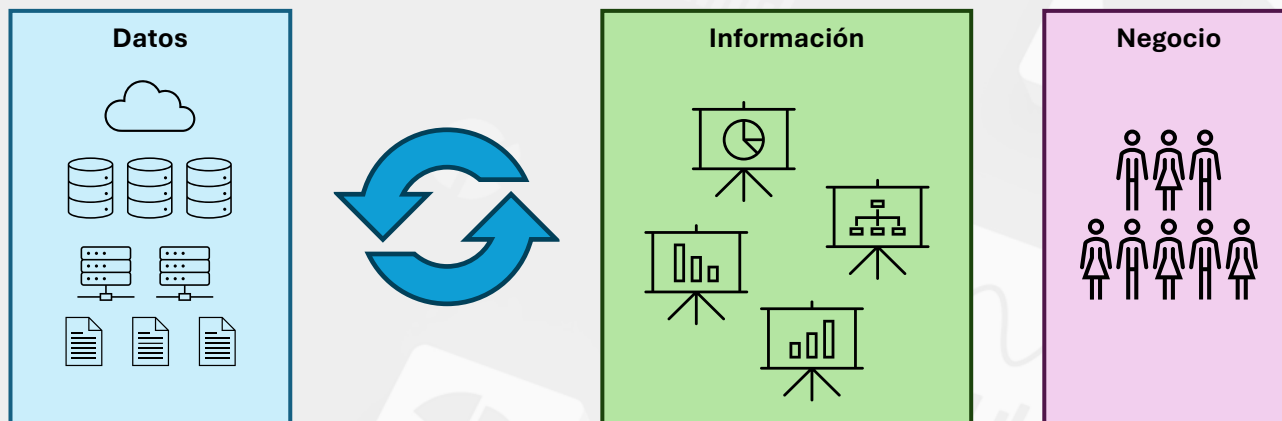
“Las plataformas de BI permiten a las empresas construir aplicaciones de BI que habilitan tres categorías de capacidades: análisis, como el OLAP; entrega de información, como en informes y cuadros de mando; e integración de plataformas, como en la gestión de metadatos de BI y en el Desarrollo de entornos”

FORRESTER

“Conjunto de metodologías, procesos, arquitecturas y tecnologías que habilitan la entrega de procesos de gestión de la información para análisis, reporting, gestión del rendimiento y entrega de información”

Business Intelligence (BI)

Es el arte de transformar los datos en información fiable y útil para el negocio facilitando el acceso y análisis por parte del negocio.



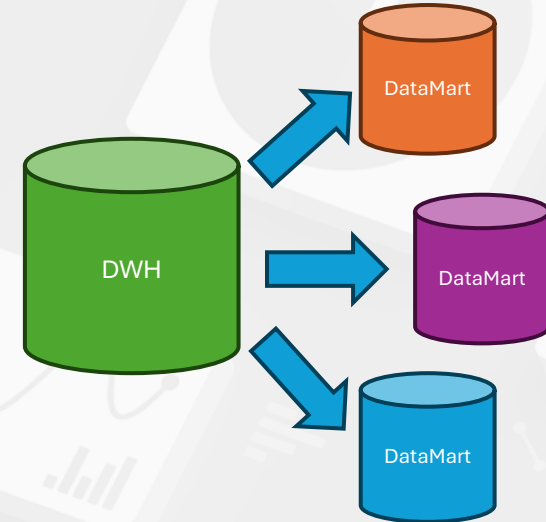
DWH (DataWarehouse)

Base de datos de soporte de toma de decisión integrada y el software asociado para recopilar, limpiar, transformar y almacenar datos.

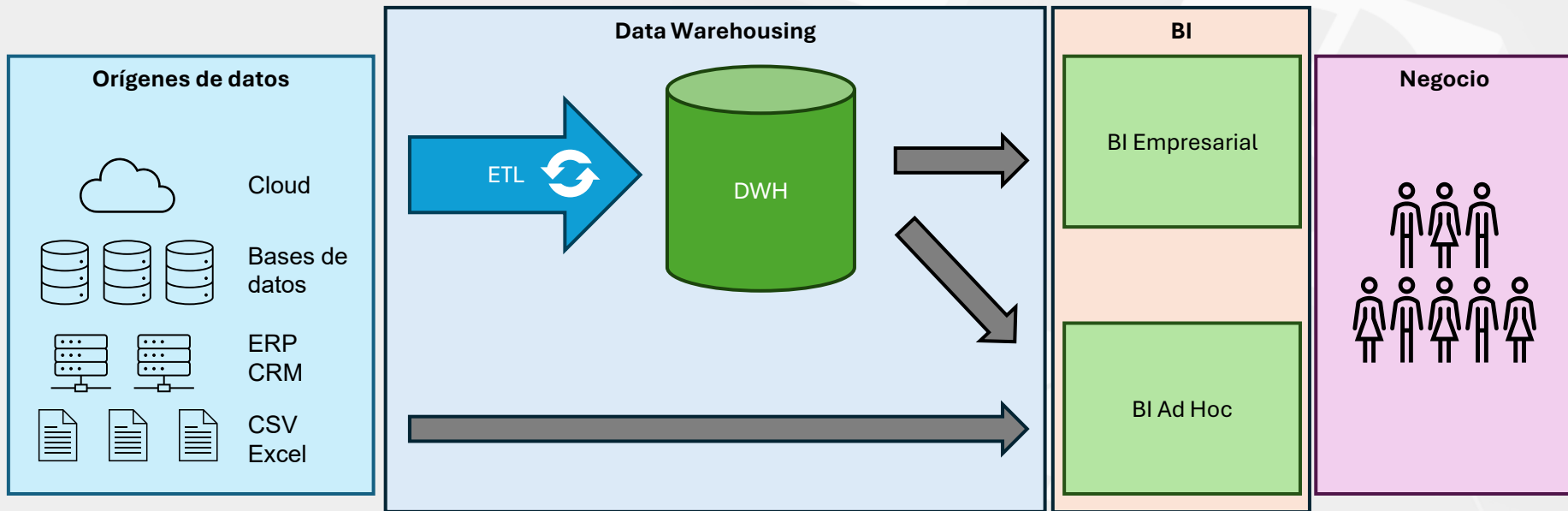
Características generales DWH:

- Enfoque global y transversal a la empresa
- Almacena datos de otros sistemas
- Organiza los datos para aumentar su valor
- Facilita el acceso y uso de los datos para el análisis
- Contiene datos integrados y fiables
- Variante en el tiempo
- No volátil

Un DataMart es un subconjunto de un DWH, un “miniDWH” con una orientación y temática específica, por ejemplo, financiera o logística.
Se utiliza cuando para proveer una solución específica y acotada a un área de la empresa.

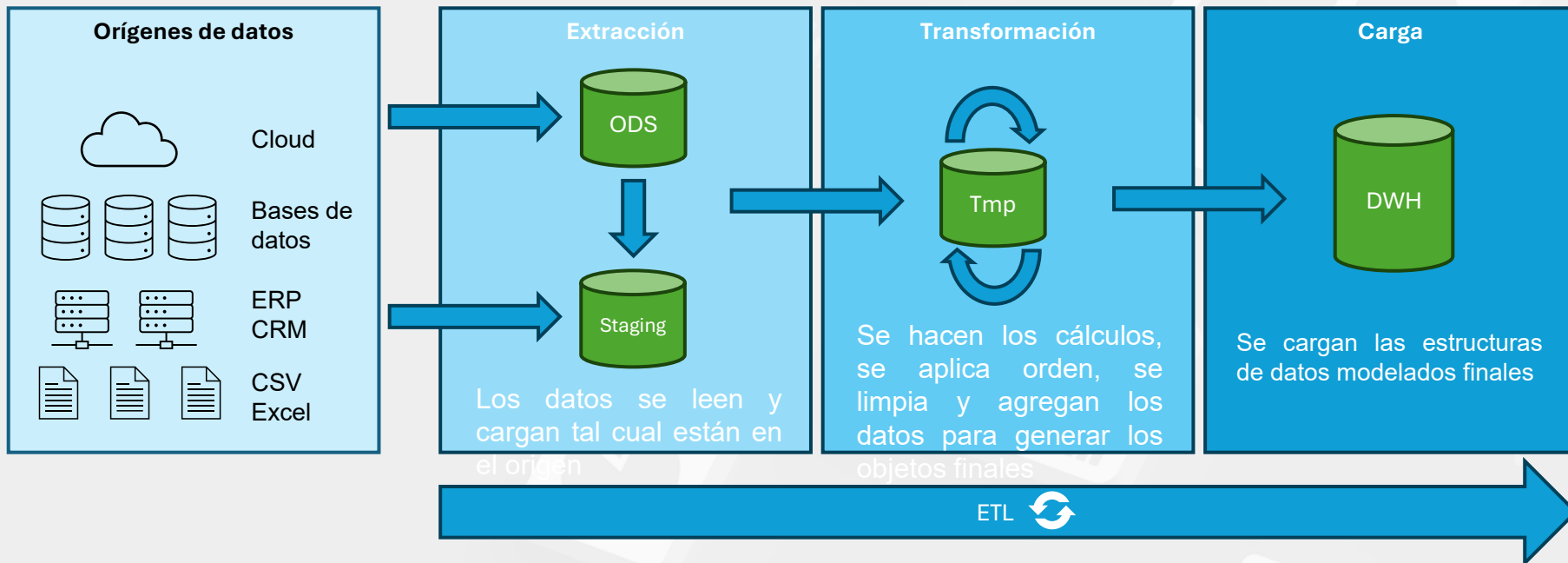


Data Warehousing y BI



ETL (Extract, Transform, Load)

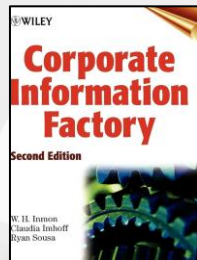
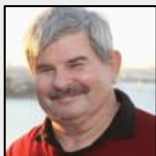
Son los procesos que extraen los datos de los orígenes, los transforman y los dejan disponibles en el DWH.



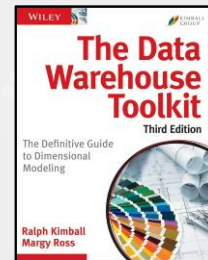
Enfoques de diseño DWH: Inmon y Kimball

- Las dos grandes corrientes de diseño de DWH fueron diseñadas por Bill Inmon y Ralph Kimball respectivamente en los 90
- Comparten características comunes, pero tienen un enfoque distinto

Inmon: Un DWH es una colección de datos orientados a un tema, integrados, variante de tiempo y no volátil para apoyar el proceso de toma de decisiones de la gerencia



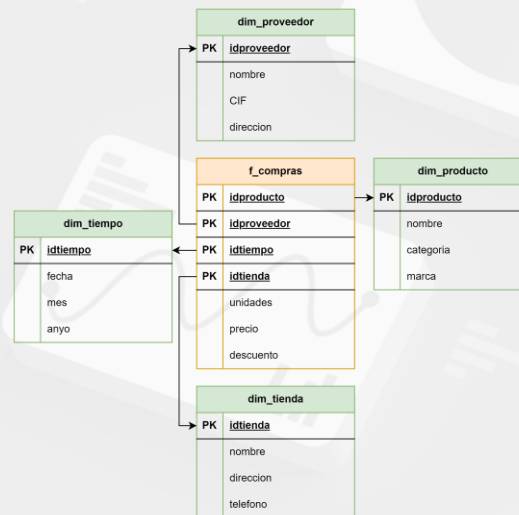
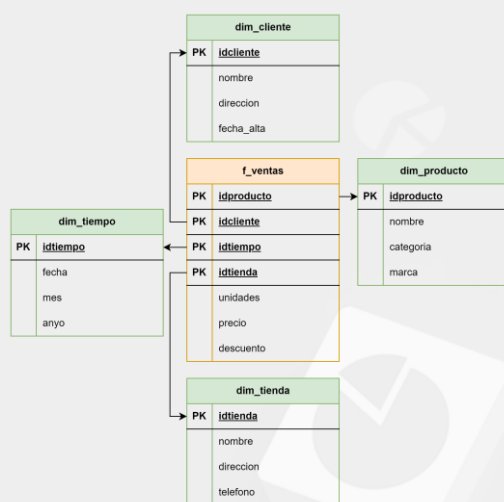
Kimball: Un DWH es una copia de los de datos de una transacción, específicamente estructurados para consultas y análisis



Enfoque Kimball: Modelo en estrella

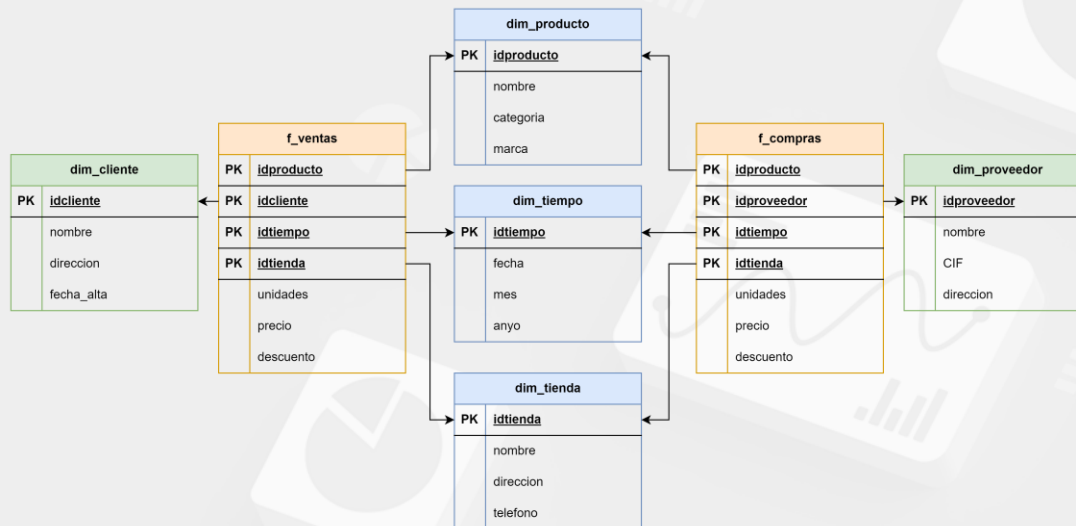
Se basa en tablas de hechos y dimensiones, una o varias estrellas componen un DataMart:

- **Tabla de hechos:** Contiene los eventos cuantificables y las medidas, datos transaccionales numéricos
- **Tabla de dimensión:** Da forma a un eje de análisis, permitiendo agregar o desglosar los datos de las tablas de hechos



Enfoque Kimball: Dimensiones conformadas

Cuando diferentes estrellas comparten dimensiones se denominan dimensiones conformadas. Estas dimensiones permiten cruzar los datos de múltiples estrellas proporcionando coherencia e interoperabilidad a lo largo del DWH. Se denomina también BUS DW al conjunto de dimensiones conformadas.



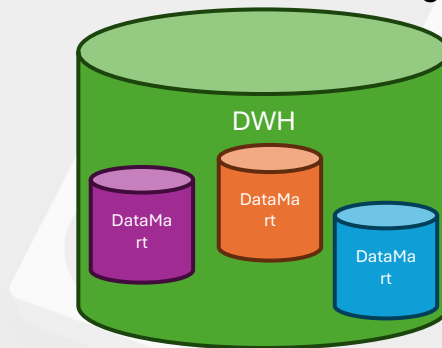
Enfoque Kimball: Ventajas e inconvenientes

Ventajas

- El modelo dimensional es rápido de construir.
- El esquema en estrella es sencillo de comprender y simplifica las consultas.
- Gracias a las dimensiones conformadas se pueden utilizar para realizar consultas a lo largo de múltiples estrellas.

Inconvenientes

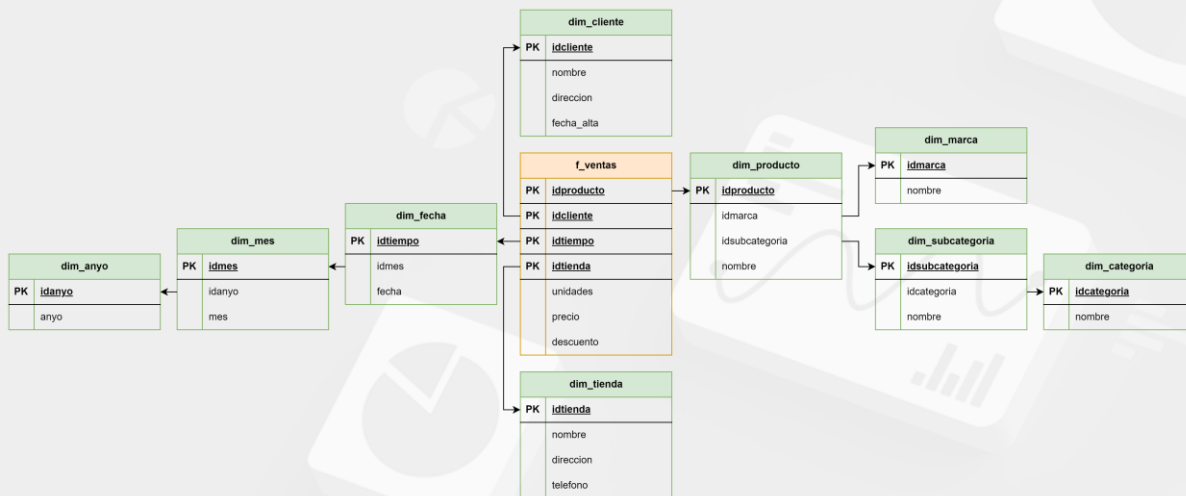
- La actualización de datos es problemática y puede generar incoherencias.
- Al estar los datos desnormalizados se pierde el concepto de única fuente de la verdad.
- Los cambios por necesidades de negocio pueden requerir de actualizaciones complejas del modelo.



Enfoque Inmon: Modelo en copo de nieve

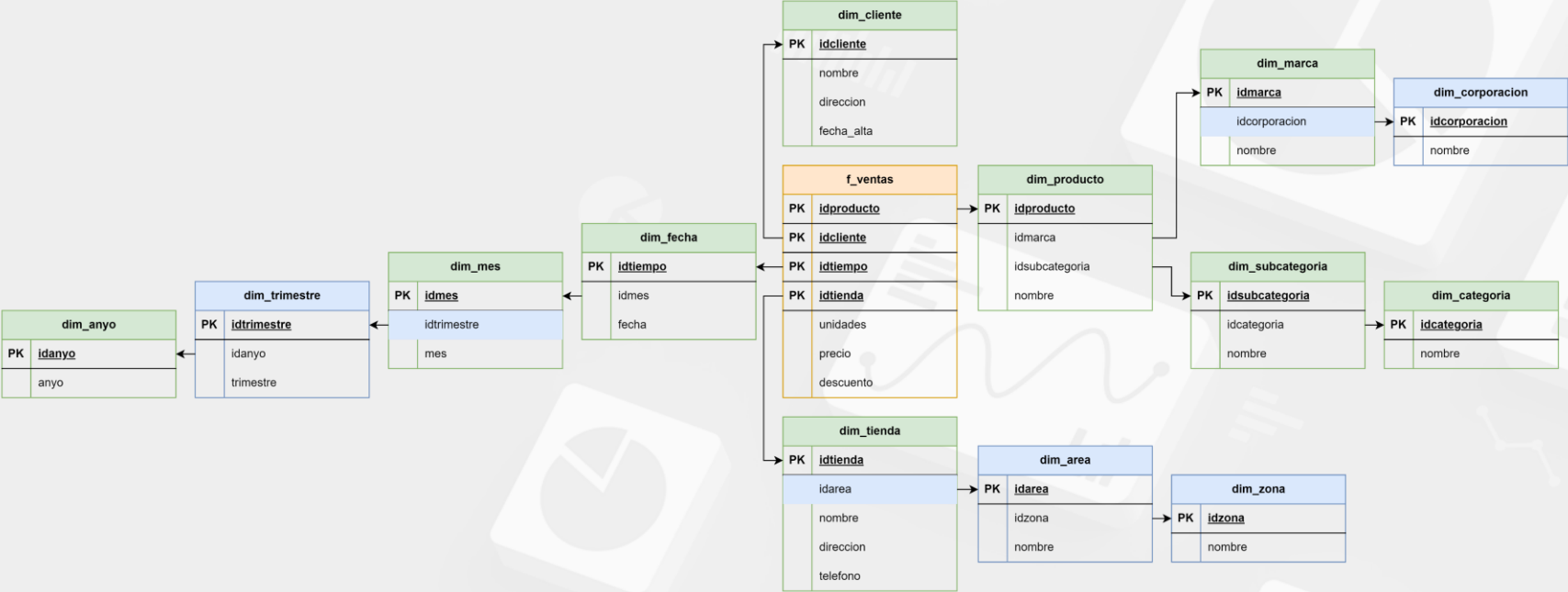
Se basa en la normalización de los datos (3NF). Tiene tablas de hechos y dimensiones como el modelo en estrella.

- **Negocio integrado en la estructura:** La propia estructura de las tablas y sus relaciones permite identificar las relaciones de las entidades de negocio y sus dependencias.
- **Cambios de negocio simples:** Aplicar cambios de negocio tales como estructuras o nuevas agrupaciones es sencillo.



Enfoque Inmon: Ejemplo de cambios

Se pueden incluir nuevos niveles en las jerarquías sin afectar los desarrollos existentes y cumpliendo con los requerimientos del negocio.



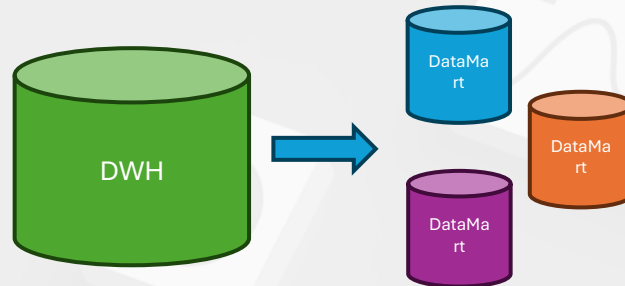
Enfoque Inmon: Ventajas e inconvenientes

Ventajas

- El DWH integrado sirve como fuente única de la verdad a nivel corporativo.
- Poca redundancia de datos al estar normalizado.
- Flexible y capaz de adoptar los cambios de negocio de forma simple.

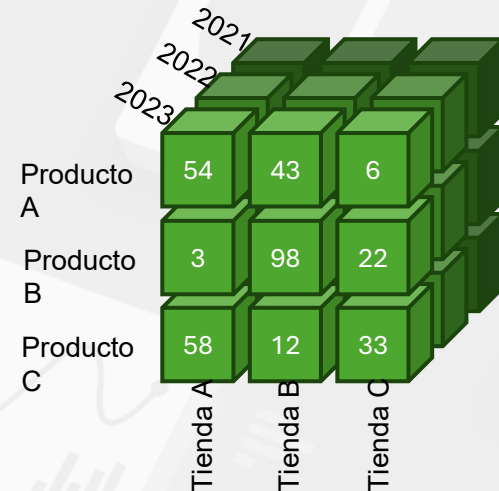
Inconvenientes

- La puesta en marcha inicial es más lenta que en el enfoque Kimball.
- El modelado requiere de una abstracción y conocimiento de negocio importante.
- Los procesos de ETL son más complejos y tienen pasos adicionales (creación de DataMarts).



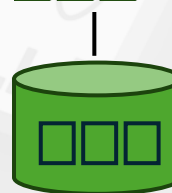
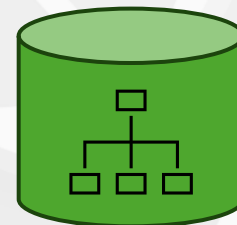
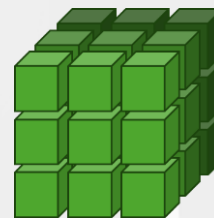
Análisis multidimensional, OLAP

- OnLine Analytical Processing
- Se basa en consultas contra dimensiones y medidas contenidas en una base de datos multidimensional, un cubo
- Cada cruce de dimensión y medida está calculado y almacenado para mejorar el rendimiento
- Suelen ser soluciones propietarias de cada fabricante



OLAP, tipos de cubo

- **MOLAP, Multidimensional OLAP:**
Basado en un formato propietario, suele ser un archivo
- **ROLAP, Relational OLAP:** Montado sobre una estructura de tablas en BBDD relacional. Se suele utilizar cuando los datos desglosados son muchos
- **HOLAP, Hybrid OLAP:**
Híbrido entre los dos anteriores, suele usar MOLAP para los agregados y ROLAP para obtener los detalles



Data Warehousing and Business Intelligence

Definition: Planning, implementation, and managing an integrated data system to support knowledge workers engaged in reporting, query, and analysis.

Goals:

1. To build and maintain the data system with the technical and business requirements needed to deliver integrated data that supports operational functions, compliance, and business intelligence.
2. To create insights to support and enable effective business analysis and decision making.

Business Drivers



Inputs:

- Business Requirements
- Scalability, Operational, Infrastructure & Support Requirements
- Data Quality, Security and Access Requirements
- IT Strategy
- Related IT Policies & Standards
- Internal Data Feeds
- Master and Reference Data
- Industry and External Data

Activities:

1. Understand Requirements (P)
2. Define and Maintain the DW and BI Architecture (P)
3. Develop the Data Warehouse and Data Marts (D)
4. Populate the Data Warehouse (D)
5. Implement the Business Intelligence Portfolio (D)
6. Maintain Data Products (O)

Deliverables:

- DW and BI Architecture
- Data Products
- Population Process
- Governance Activities
- Data Lineage
- Training and Adoption Plan
- Release Plan
- Production Support Process
- Load Tuning Activities
- BI Activity Monitoring
- Reporting Strategy

Suppliers:

- Business Executive
- Governance Body
- Enterprise Architecture
- Data Producers
- Information Consumers
- Subject Matter Experts

Participants:

- Sponsors & Product Owner
- Architects and Analysts
- DW/BI Specialists (BI Platform, Data Storage)
- Project Manager
- Change Manager

Consumers:

- Information Consumers
- Customers
- Partners
- Managers and Executives
- Subject Matter Experts

Technical Drivers



Techniques:

- Prototypes to Drive Requirements
- Self-Service BI
- Queryable Audit Data

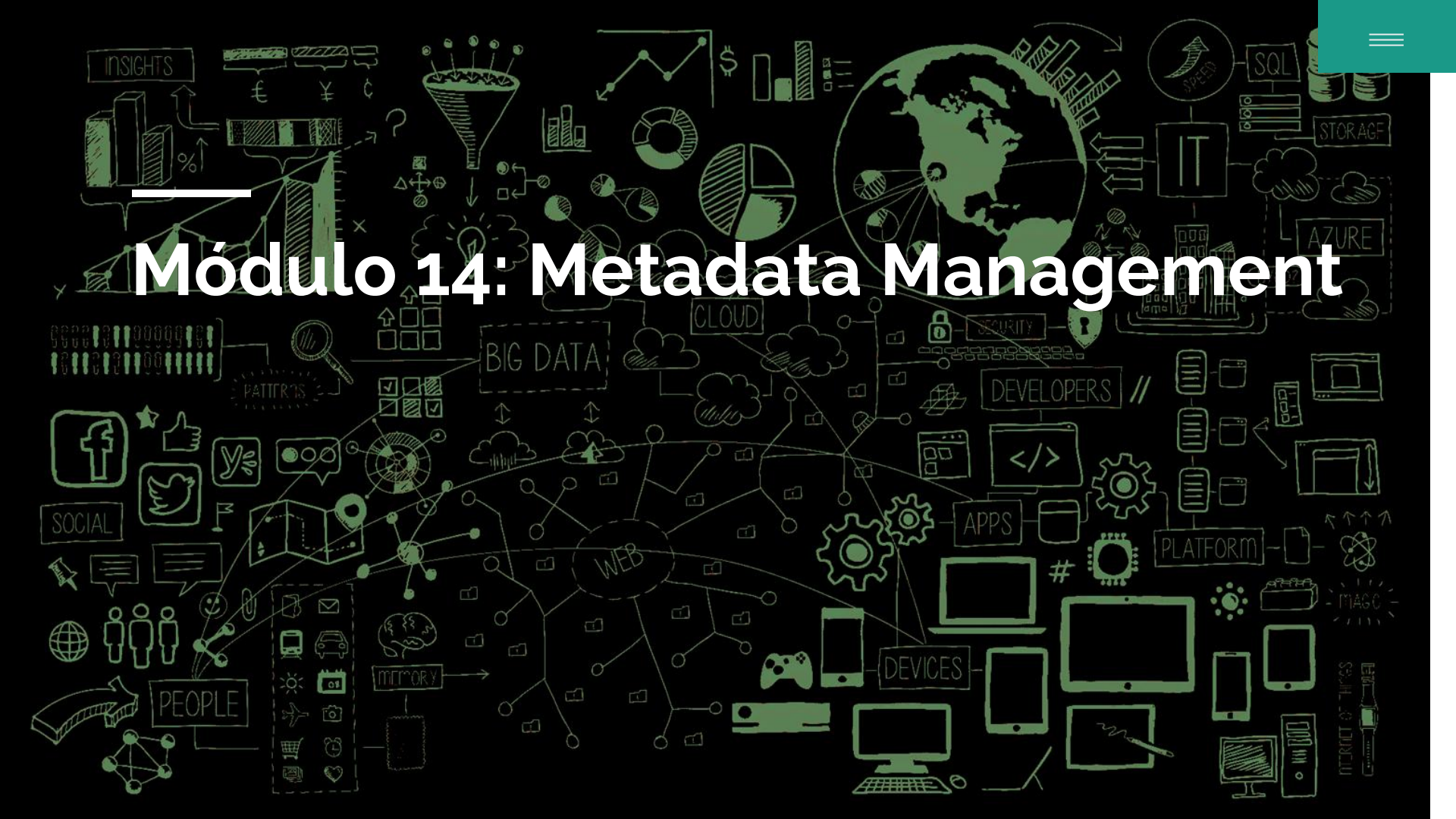
Tools:

- Metadata Repositories
- Data Integration Tools
- Business Intelligence Tools

Metrics:

- Usage Metrics
- Subject Area Coverage Percentages
- Response/Performance Metrics

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



Módulo 14: Metadata Management

Metadatos: Tipología de Metadatos

Los metadatos suelen clasificarse en tres tipos: de negocios, técnicos y operativos.

Metadatos de Negocio

- Los metadatos de negocios se centran en gran medida en el contenido y la condición de los datos e incluyen detalles relacionados con la gobernanza de los datos.
- Incluyen los nombres y definiciones no técnicos de conceptos, áreas temáticas, entidades y atributos; tipos de datos de atributos y otras propiedades de atributos; descripciones de rangos; cálculos; algoritmos y reglas de negocios; valores de dominio válidos y sus definiciones.

Metadatos Técnicos

- Los metadatos técnicos proporcionan información sobre los detalles técnicos de los datos, los sistemas que los almacenan y los procesos que los mueven dentro de los sistemas y entre ellos.

Metadatos Operacionales

- Metadatos operacionales describen los detalles del procesamiento y el acceso a los datos.

Business Glossary

El propósito de un glosario de negocios (business glossary) es documentar y almacenar los conceptos y la terminología comercial de una organización, así como las definiciones y las relaciones entre esos términos.

La aplicación del glosario de negocios está estructurada para satisfacer los requisitos funcionales de las tres audiencias principales:

- **Usuarios de negocios:** Los analistas de datos, los analistas de investigación, administración y el personal ejecutivo utilizan el glosario empresarial para comprender la terminología y los datos.
- **Data Steward:** los *Data Stewards* utilizan el glosario empresarial para gestionar el ciclo de vida de los términos y definiciones y para mejorar el conocimiento de la empresa asociando los activos de datos con los términos del glosario; por ejemplo, vinculando los términos a las métricas empresariales, los informes, el análisis de la calidad de los datos o los componentes tecnológicos. Los administradores de datos plantean cuestiones de terminología y uso y ayudan a resolver las diferencias en toda la organización.
- **Usuarios técnicos:** Los usuarios técnicos utilizan el glosario empresarial para tomar decisiones sobre arquitectura, diseño de sistemas y desarrollo, y para realizar análisis de impacto.

Data Catalog (Data Marketplace)

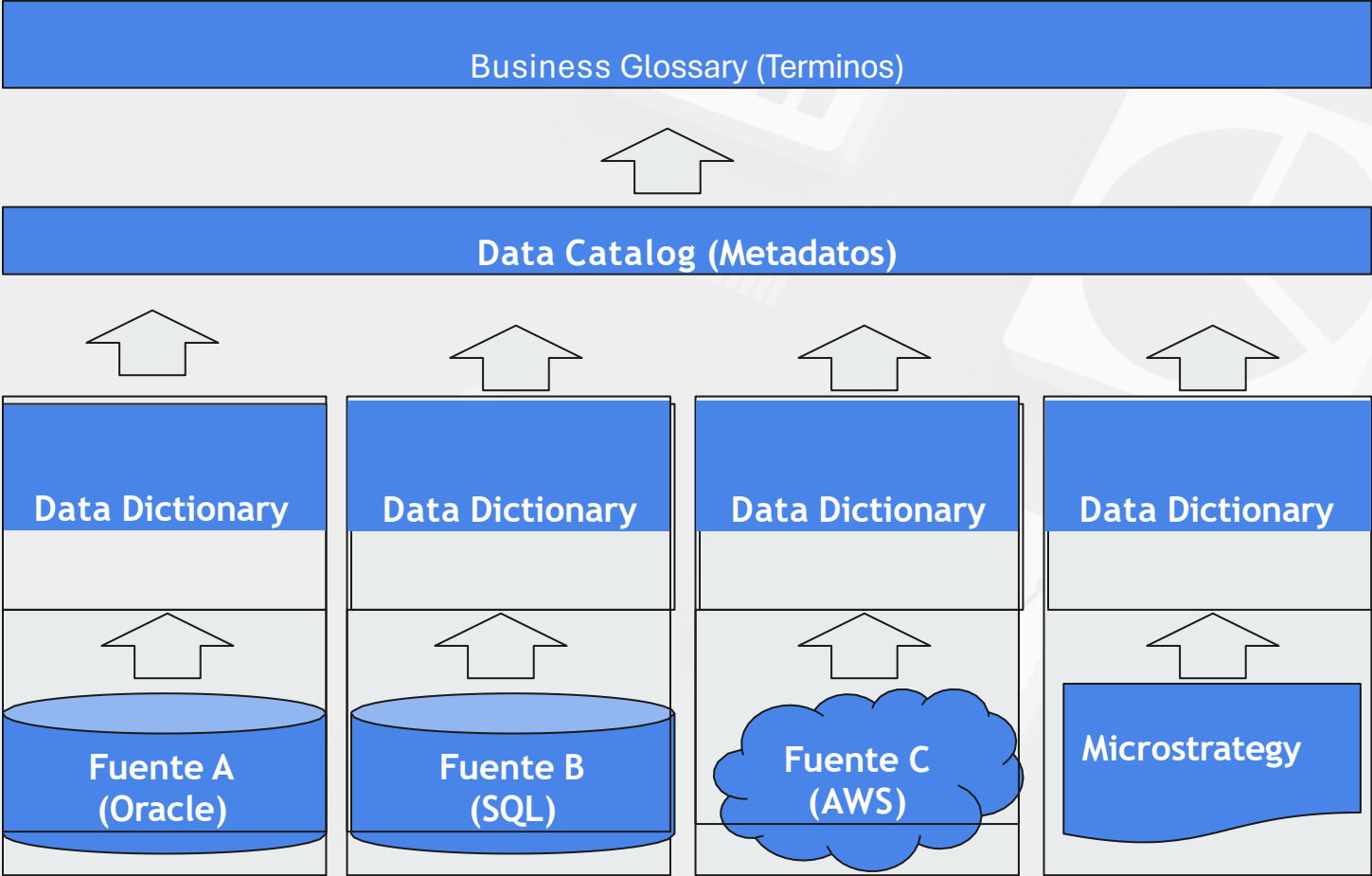
El catálogo de datos sirve como un directorio de un solo punto para localizar información y además proporciona el mapeo entre el glosario de negocios y los diccionarios de datos.

- **El catálogo de datos es un activo de toda la empresa** que proporciona una única fuente de referencia para la localización de cualquier conjunto de datos necesarios para diversas necesidades como las operacionales, BI, analíticas, de ciencia de los datos, etc.
- Si una división de una empresa es significativamente diferente de otras, sería razonable que el catálogo de datos fuera exclusivo de la división y no de la empresa.
- Normalmente el catálogo de datos se desarrolla después de la creación satisfactoria tanto del Business Glossary como de los diccionarios de datos, pero también puede ensamblarse de forma incremental a medida que los otros dos activos evolucionen con el tiempo.

Data Dictionary (Diccionario de datos)

Un diccionario de datos debe centrarse en las descripciones y detalles que conlleva el almacenamiento de datos está directamente ligado a los metadatos.

- Debería haber un diccionario de datos para **cada fuente de datos** de la empresa.
- El diccionario de datos incluye detalles sobre los datos como el tipo de datos, la longitud permitida, el linaje, las transformaciones, etc.
- Estos metadatos ayudan a los **arquitectos, ingenieros y científicos de datos** a entender cómo unir, consultar y reportar los datos, y también explica la granularidad.
- Debido a la necesidad de conocimientos técnicos y metadatos, la responsabilidad de la propiedad de un diccionario de datos recae en la tecnología, frecuentemente con funciones como administradores de bases de datos, ingenieros de datos, arquitectos de datos y/o administradores de datos.



Herramientas de gestión de configuración

- Las **CMDB** (herramientas o bases de datos de gestión de la configuración) proporcionan la capacidad de gestionar y mantener metadatos específicamente relacionados con los activos de TI, las relaciones entre ellos y los detalles contractuales de los activos.
- Cada elemento de la base de datos CMDB se denomina elemento de configuración (CI = Configuration Item). Los Metadatos estándares se recopilan y administran por cada tipo de CI.
- Los repositorios proporcionan mecanismos para vincular los activos en el repositorio de Metadatos con los detalle físicos reales de la implementación en CMDB para dar una imagen completa de los datos y las plataformas.

Metadata Management

Definition: Planning, Implementation, and control activities that contributes to the ability to process, maintain, integrate, secure, audit and govern other data

Goals:

1. Provide organizational understanding of business terms and usage including technical lineage.
2. Collect and integrate metadata from diverse sources.
3. Provide a standard way to access metadata and enable known level of trust in data exchange.
4. Ensure metadata quality, consistency, currency, and security.

Business Drivers

Inputs:

- Business and Technical Metadata
- Operational Requirements
- Metadata Policy and Standards
- Data Architecture
- Metadata Issues
- Process Metadata
- Operational Metadata
- Data Governance Metadata

Activities:

1. **Define Metadata Strategy (P)**
2. **Understand Metadata Requirements (P)**
 1. Business User Requirements
 2. Technical User Requirements
3. **Define Metadata Architecture (P)**
 1. Create Metamodel (D)
 2. Apply Metadata Standards (C)
 3. Manage Metadata Stores (C)
4. **Create and Maintain Metadata (O)**
 1. Integrate Metadata (O)
 2. Distribute and Deliver Metadata (O)
5. **Query, Report and Analyze Metadata (O)**

Deliverables:

- Metadata Strategy
- Metadata Standards
- Metadata Architecture
- Metamodel
- Unified Metadata
- Metadata Stores
- Data Lineage
- Impact Analysis
- Metadata Control Process

Suppliers:

- Business Data Stewards
- Data Managers
- Data Governance Bodies
- Data Modelers
- Database Administrators

Participants:

- Data Stewards
- Project Managers
- Data Architects
- Business Analysts
- System Analysts

Consumers:

- Application Developers
- Data Analyst
- Data Integrators
- Business Users
- Knowledge Workers
- Customers & Collaborators
- Data Scientists
- Data Journalists
- Data Curators

Technical Drivers

Techniques:

- Data Lineage and Impact Analysis
- Metadata for Big Data Ingest

Tools:

- Metadata Repository Management Tools
- Metadata Repositories in other Tools
- Integration Tools

Metrics:

- Metadata Coverage Scorecard
- Metadata Quality Scorecard
- Metadata Repository Activities

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

¡Gracias!



GOLD SPONSORS



SILVER SPONSORS



Entidades académicas

