

UNIDAD 3

Explotación de Datos

Explotación de Datos

Contenido de la Unidad

- 1. Introducción a Data Lakes**
2. Técnicas de DM y descubrimiento de conocimiento
3. Business Intelligence (BI)
4. Motivación y aplicabilidad de BI
5. Curado de datos
6. Tableros de Control
7. Práctica: Diseño de Dashboards y Tableros
8. Resumen

Explotación de Datos

Introducción a Data Lakes - DW

Un **Datawarehouse** es una gran colección de datos de negocio procedentes de un gran número de **fuentes distintas** reunidos en un **único almacén de datos homogéneo**.

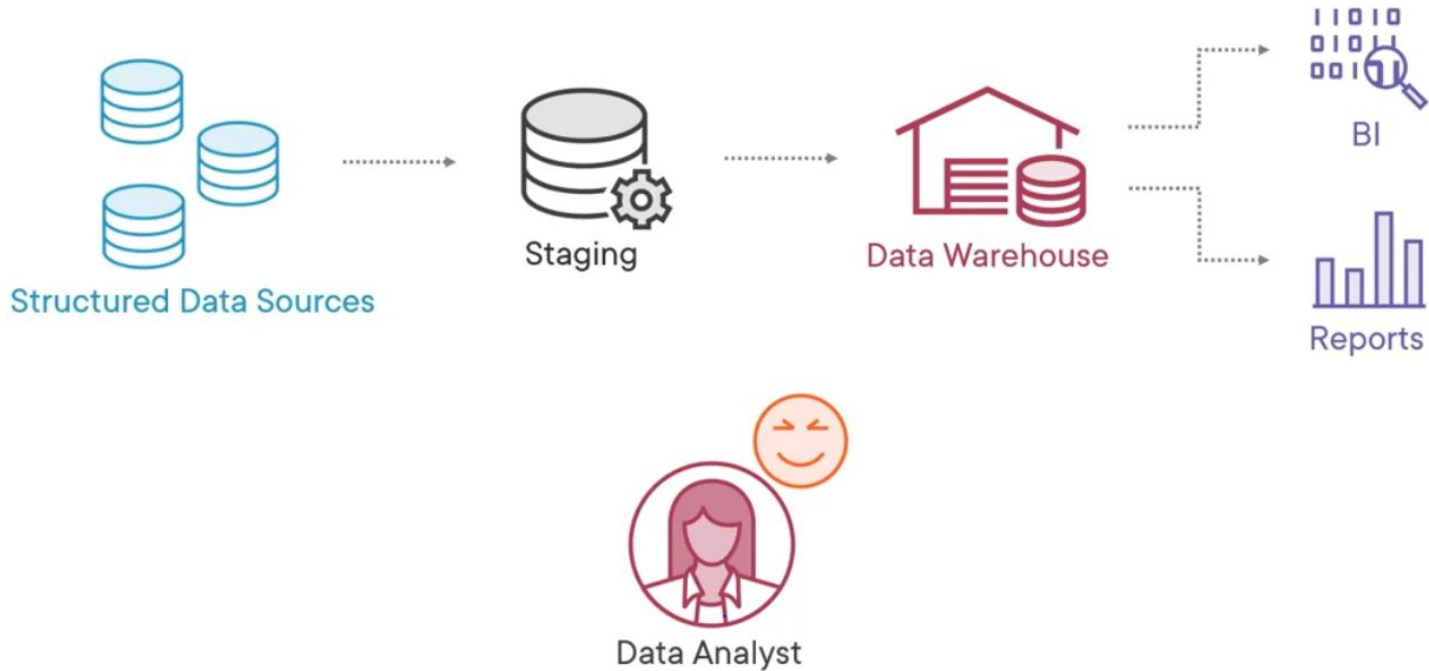
Estas plataformas están específicamente diseñadas para realizar análisis de grandes cantidades de **datos estructurados**.

Por lo general, un sistema de almacén de datos extrae los datos de diversos **sistemas transaccionales (OLTP)**, para después limpiar e importar esos datos con la finalidad de hacerlos coincidir con los formatos y estándares presentes en el almacén de datos.

Esto permite que los datos se almacenen en archivos o carpetas organizados para que estén disponibles inmediatamente para la elaboración de informes o el análisis de datos.

Explotación de Datos

Introducción a Data Lakes - DW



Explotación de Datos

Introducción a Data Lakes - Data Lake

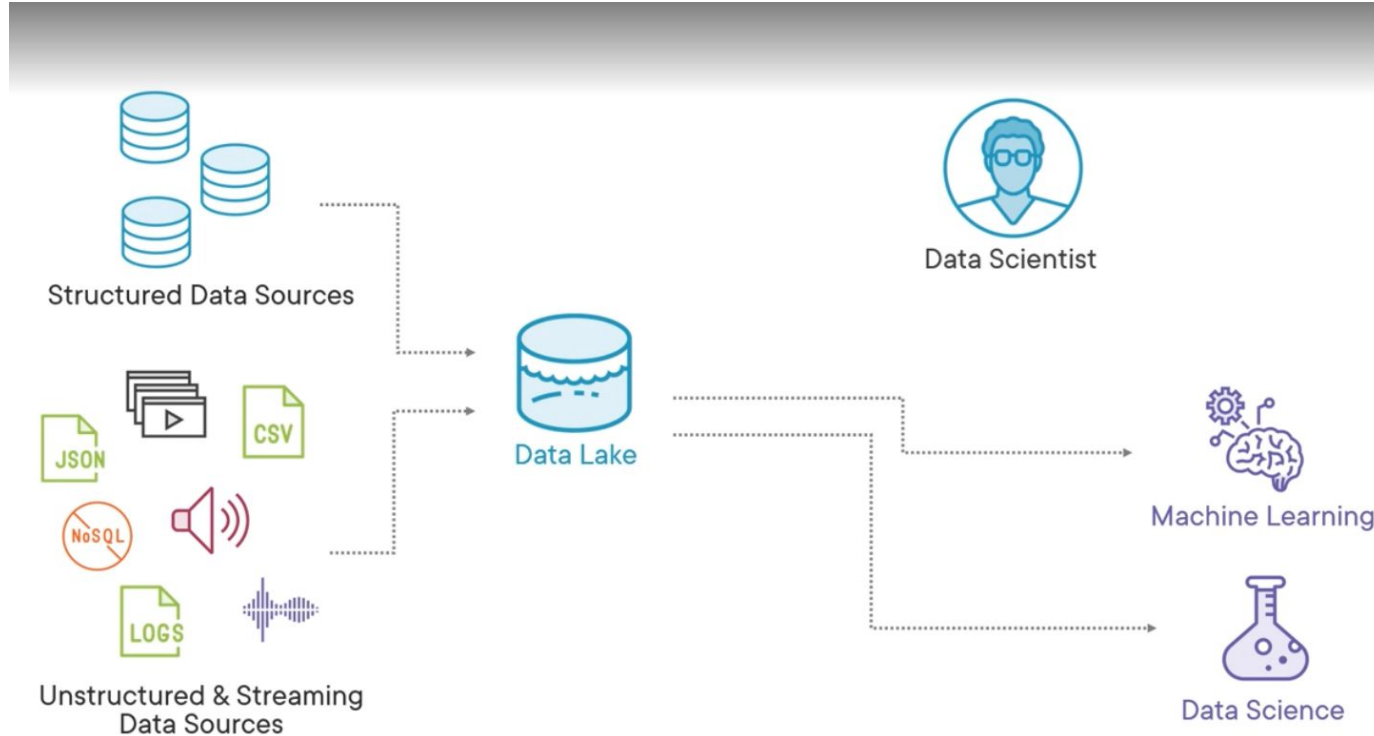
Un lago de datos (o data lake) almacena en su **formato nativo y a escala todo tipo de datos**, ya sean **sin procesar, estructurados o desestructurados**, procedentes de **todas las fuentes de datos de la empresa**.

Los datos se añaden al lago de datos como están, es decir, los nuevos datos no se formatean para equipararse al resto de datos ya existentes en el sistema.

Los lagos de datos desempeñan un papel fundamental permitiendo que los datos estén disponibles para sistemas de **IA y ML y análisis de Big Data**.

Explotación de Datos

Introducción a Data Lakes - Data Lake



Explotación de Datos

Introducción a Data Lakes - Data Lake House

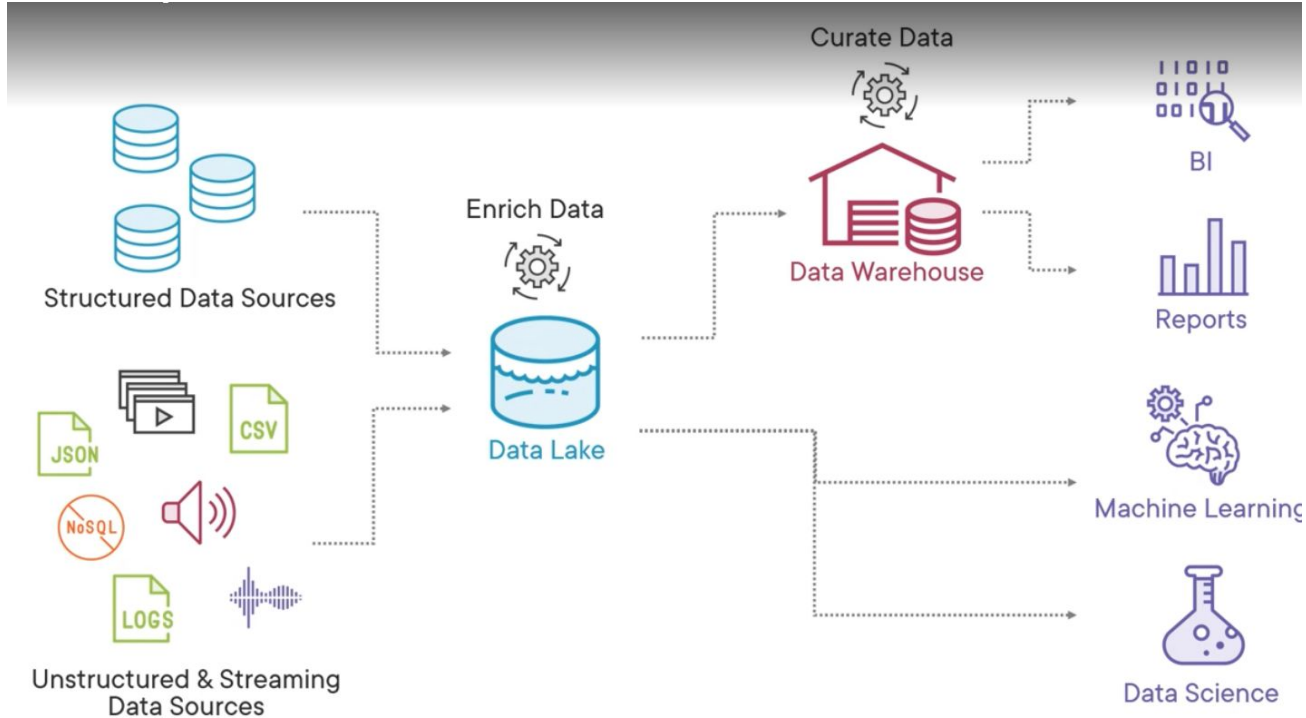
Un **Data Lake House** es una arquitectura nueva y abierta que combina las ventajas de flexibilidad y escalabilidad de un **lago de datos con las estructuras de datos y características de gestión de datos similares de un datawarehouse**.

Esta combinación de características hace posible que los equipos de data science dispongan de agilidad, ya que así son capaces de usar los datos sin tener que acceder a varios sistemas.

Los Data Lake Houses también garantizan que los científicos de datos dispongan de los datos más completos y actualizados.

Explotación de Datos

Introducción a Data Lakes - Data Lake House



Explotación de Datos

Introducción a Data Lakes

Dudas / Preguntas



Explotación de Datos

Contenido de la Unidad

1. **Introducción a Data Lakes**
2. **Técnicas de DM y descubrimiento de conocimiento**
3. Business Intelligence (BI)
4. Motivación y aplicabilidad de BI
5. Curado de datos
6. Tableros de Control
7. Práctica: Diseño de Dashboards y Tableros
8. Resumen

Explotación de Datos

Técnicas de DM y descubrimiento de conocimiento - Parte 1

Las herramientas de Data Mining (DM) predicen tendencias y comportamientos que ayudan a descubrir información relevante a través de técnicas que buscan patrones ocultos

Explotación de Datos

Descubrimiento de conocimiento - Data Mining

El proceso de DM se puede definir mediante los siguientes seis pasos básicos:

1. [Definición del problema](#)
2. [Preparación de datos](#)
3. [Explorar los datos](#)
4. [Creación de modelos](#)
5. [Exploración y validación de modelos](#)
6. [Implementar y actualizar los modelos](#)

Explotación de Datos

Técnicas de DM y descubrimiento de conocimiento - Parte 1

Redes neuronales artificiales

Radica en el aprendizaje secuencial, el hecho de utilizar transformaciones de las variables originales para la **predicción** y la no linealidad del modelo.

Árboles de decisión

Permiten obtener de forma visual las reglas de decisión bajo las cuales operan los consumidores, a partir de datos históricos almacenados.

Agrupamiento (Clustering)

Agrupar un conjunto de observaciones en un número dado de clusters o grupos. Está basado en la idea de similitud de los grupos.

Explotación de Datos

Técnicas de DM y descubrimiento de conocimiento - Parte 2

Algoritmo Jerárquico

Se debe calcular la distancia entre los pares de objetos o clusters; se buscan los dos clusters más cercanos, éstos se juntan y constituyen uno solo. Se repiten los pasos hasta que no queden pares de comparación.

Regla de Inducción

Consiste en derivar un conjunto de reglas para clasificar casos, generan un conjunto de reglas independientes que permiten contrastar árboles de decisión y patrones a partir de los datos de entrada.

La información de entrada será un conjunto de casos en que se ha asociado una clasificación a un conjunto de variables o atributos.

Explotación de Datos

Técnicas de DM y descubrimiento de conocimiento

Dudas / Preguntas



Explotación de Datos

Contenido de la Unidad

1. Introducción a Data Lakes
2. Técnicas de DM y descubrimiento de conocimiento
3. **Business Intelligence (BI)**
4. Motivación y aplicabilidad de BI
5. Curado de datos
6. Tableros de Control
7. Práctica: Diseño de Dashboards y Tableros
8. Resumen

Explotación de Datos

Business Intelligence - Conceptos

“Business Intelligence suele definirse como la **transformación de los datos** de la compañía en **conocimiento** para obtener una ventaja competitiva” (Fuente: Gartner Group).

- Por un lado: se refiere al tipo de análisis de datos que apunta el conocimiento de las actividades y oportunidades de la empresa. Haciendo las preguntas correctas sobre la base de datos propios, se pueden tomar decisiones más acertadas para alcanzar objetivos empresariales.
- Por otro lado: se refiere al conjunto de tecnologías que soportan este tipo de análisis de datos. Como una evolución de los DSS, las herramientas de BI permiten consultas, minería de datos, análisis estadístico, reporting, modelo de escenarios, visualización de datos, tableros de comando

Explotación de Datos

Business Intelligence - Conceptos

Business Intelligence

**Utilizar el activo que representan los datos
para tomar mejores decisiones de negocio.**

Explotación de Datos

Business Intelligence - Definiciones

Definiciones

Gartner: Es un término general que incluye las aplicaciones, la infraestructura y las herramientas, y las mejores prácticas que permiten el acceso y el análisis de la información para mejorar y optimizar las decisiones y el rendimiento.

Business Intelligence: Es un término que se refiere a ideas, prácticas y tecnologías para convertir los datos sin procesar en información que las empresas pueden utilizar para tomar mejores decisiones organizacionales.

Explotación de Datos

Business Intelligence - Definiciones



Definiciones

Forrester: Conjunto de metodologías, procesos, arquitecturas y tecnologías que aprovechan el rendimiento de los procesos de gestión de la información para el análisis, la generación de informes, la gestión del rendimiento y la entrega de información. La cobertura de investigación incluye tableros ejecutivos así como herramientas de consulta e informes.

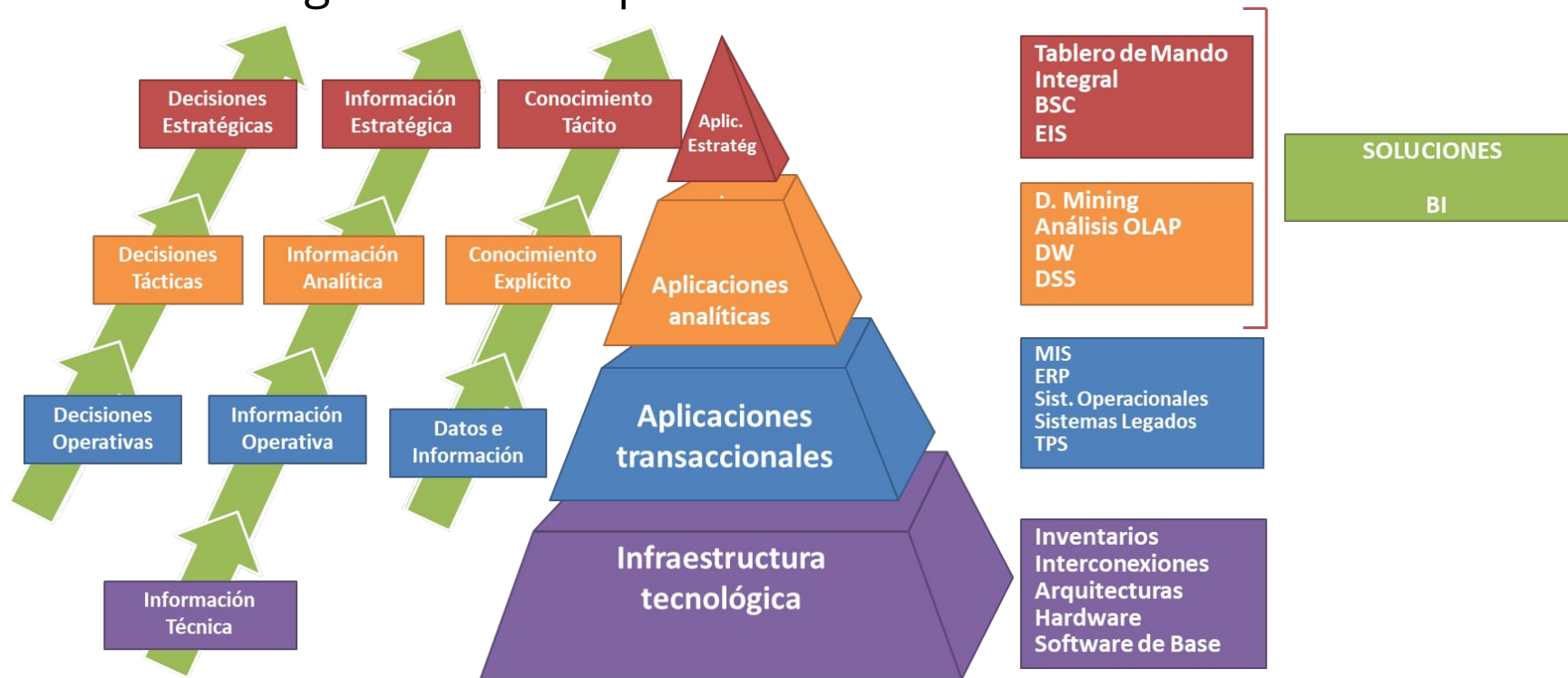
Gartner: <https://www.gartner.com/it-glossary/business-intelligence-bi/>

BusinessIntelligence: <https://businessintelligence.com/>

Forrester: <https://www.forrester.com/Business-Intelligence>

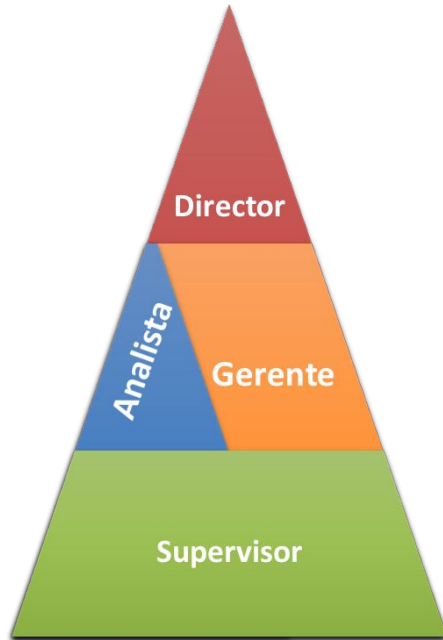
Explotación de Datos

Business Intelligence - Conceptos



Explotación de Datos

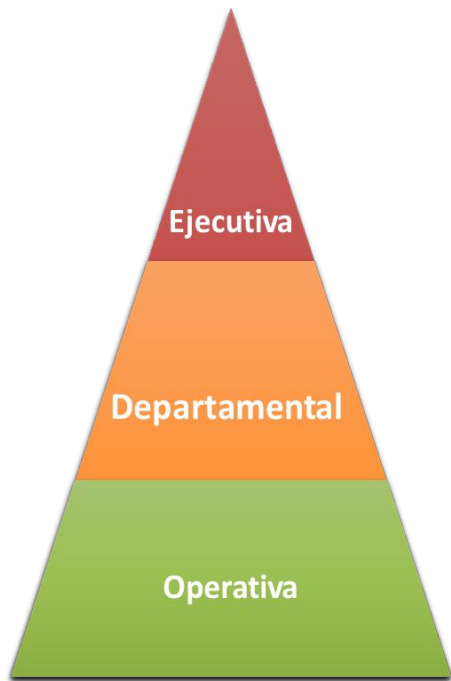
Business Intelligence - Necesidades según Rol



Características
<ul style="list-style-type: none"> ▼ Indicadores ▼ Semáforos ▼ Alertas ▼ Pocos Números – Significativos
<ul style="list-style-type: none"> ▼ Iteración fácil ▼ Libertades estilo Tabla dinámica ▼ Posibilidad de Operaciones sencillas ▼ Análisis multidimensional ▼ Publicación de resultados al resto de la organización
<ul style="list-style-type: none"> ▼ Acceso a Todos los datos disponibles ▼ Libertad de análisis
<ul style="list-style-type: none"> ▼ Reportes Pre-Definidos y pre calculados ▼ Acceso con restricciones

Explotación de Datos

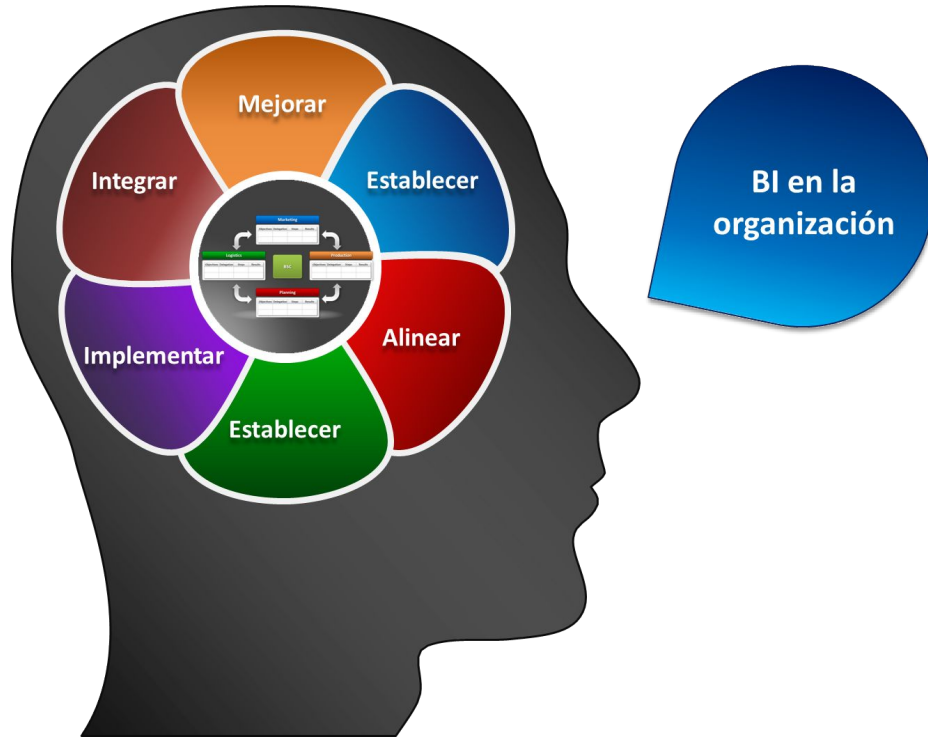
Business Intelligence - Necesidades según Rol



Característica	Deseable
<ul style="list-style-type: none"> • Todas las Áreas. • Poco detalle. 	<ul style="list-style-type: none"> • Solo por Excepción. • En cualquier lugar.
<ul style="list-style-type: none"> • Foco en su área. • Reportes de Evolución. • Análisis a largo plazo. • Consultas libres. 	<ul style="list-style-type: none"> • Historia completa y de libre acceso
<ul style="list-style-type: none"> • Minuto a Minuto. • Poca Historia. • Consultas pre-armadas. 	<ul style="list-style-type: none"> • Casi On-Line • Cubre Falencias de los sistemas operacionales

Explotación de Datos

Business Intelligence - Necesidades según Rol



Lo que permite al comandante sabio golpear, conquistar, y lograr cosas que están fuera del alcance de los hombres ordinarios, es el conocimiento previo.

Sun Tzu

Explotación de Datos

Business Intelligence - Necesidades según Rol

1	Definiciones	7	Descripción del ecosistema
2	Usos del BI	8	Problemáticas de BI
3	Interpretaciones	9	Lo que se quiere lograr
4	Tecnología propuesta	10	Enfoque de la transformación
5	Diferencias BI vs AN	11	Beneficios esperados
6	Ecosistema de BI	12	Modelo de madurez

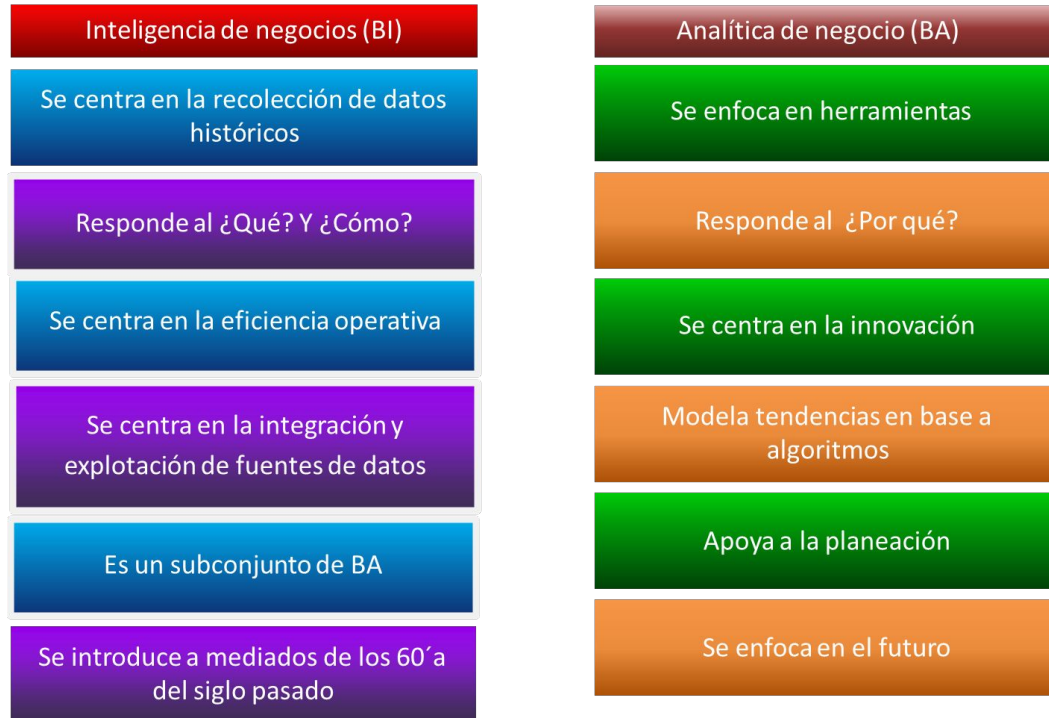
Explotación de Datos

Business Intelligence - Tipos de Uso



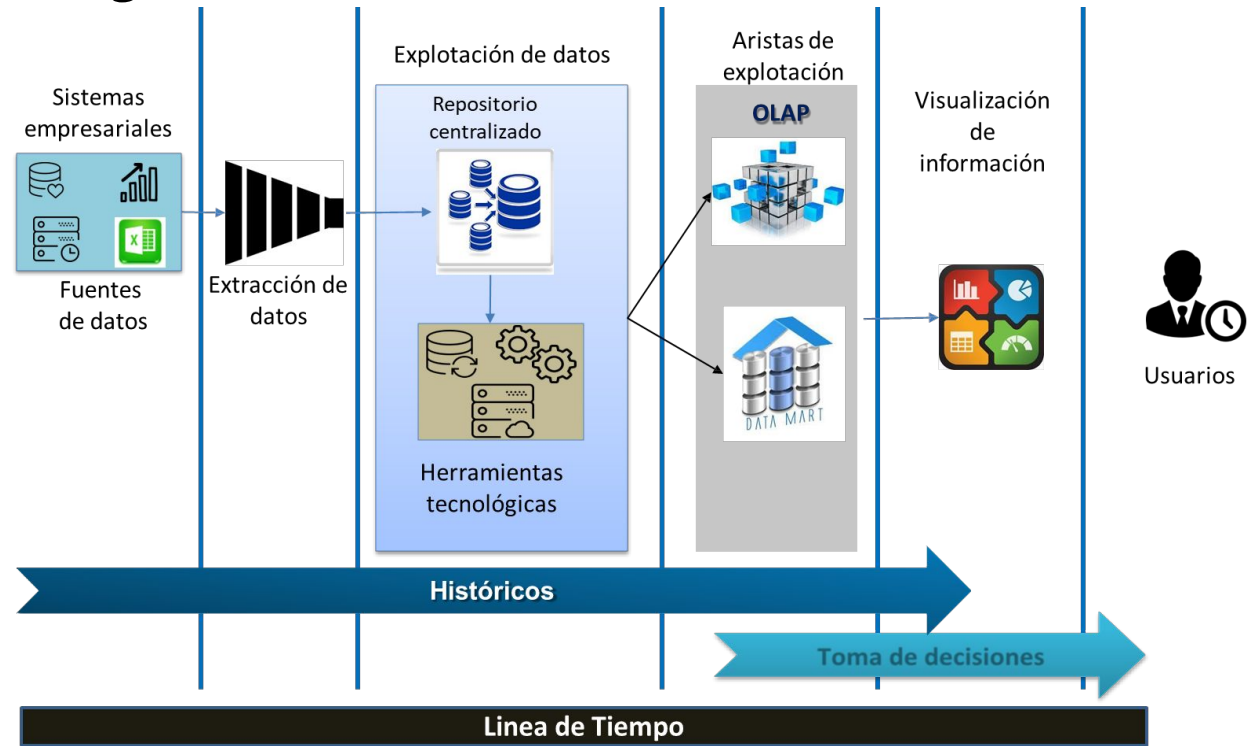
Explotación de Datos

Business Intelligence - Diferencias entre BI y AB



Explotación de Datos

Business Intelligence - Ecosistema de BI



Explotación de Datos

Business Intelligence - Ecosistema de BI - Sistemas Empresariales



- Son Sistemas de Información con los datos de la **operación diaria**.
- Los **datos** están organizados y relacionados para facilitar operaciones de lectura y escritura
- Normalmente son consultados mediante **SQL**.
- Las operaciones realizadas sobre estos sistemas se conocen como **“transacciones”**
- Estos sistemas son soportados y administrados por **personal especializado de TI**.

Explotación de Datos

Business Intelligence - Ecosistema de BI - Extracción y Filtrado



- Procesos que permiten integrar los datos de las distintas fuentes de información.
- Facilita el movimiento de datos desde las fuentes al destino.
- Permite formatear y limpiar los datos en base a reglas
- Es el preámbulo para comenzar a analizar los datos
- Permite preparar el escenario para la toma de decisiones.

Explotación de Datos

Business Intelligence - Ecosistema de BI - Explotación



- La explotación se realiza en un repositorio central donde se ejecutarán las reglas de negocio.
- Son sistemas con datos granulares organizados en dimensiones para su mejor explotación.
- Plataformas tecnológicas con motores que facilitan la explotación de datos.
- Organización en tablas de hechos y dimensiones.
- Pueden ser columnares o con motores que emulan sistemas de información columnares.

Explotación de Datos

Business Intelligence - Ecosistema de BI - Aristas de Explotación



- Permiten ver diferentes vistas de los datos.
- Pueden manejarse por área de negocio (DataMart) o por diferentes aristas (cubos) para distintas áreas.
- Son sistemas especializados en reglas de negocios y muy granulares.
- Son sistemas con alta transaccionalidad.
- Permiten modelar tanto escenarios puntuales como amplios para la toma de decisiones.

Explotación de Datos

Business Intelligence - Ecosistema de BI - Visualización



- Permiten mostrar la información en forma de tableros de control.
- Representan los datos en forma de información.
- Pueden segmentarse por tipo de usuarios.
- Permiten dar visibilidad a la toma de decisiones.
- Pueden representar tanto datos operativos como estratégicos.

Explotación de Datos

Business Intelligence (BI)

Dudas / Preguntas








Explotación de Datos

Contenido de la Unidad


1. Introducción a Data Lakes
2. Técnicas de DM y descubrimiento de conocimiento
3. Business Intelligence (BI)
- 4. Motivación y aplicabilidad de BI**
5. Curado de datos
6. Tableros de Control
7. Práctica: Diseño de Dashboards y Tableros
8. Resumen

Explotación de Datos

Motivación y aplicabilidad de BI



Interna, con poca fluidez.
Documentos no integrados y alineados.
Silos de Información en diferentes archivos.
No optimizadas para el desarrollo de propuestas.
Sin visibilidad integral de información para la toma de decisiones.



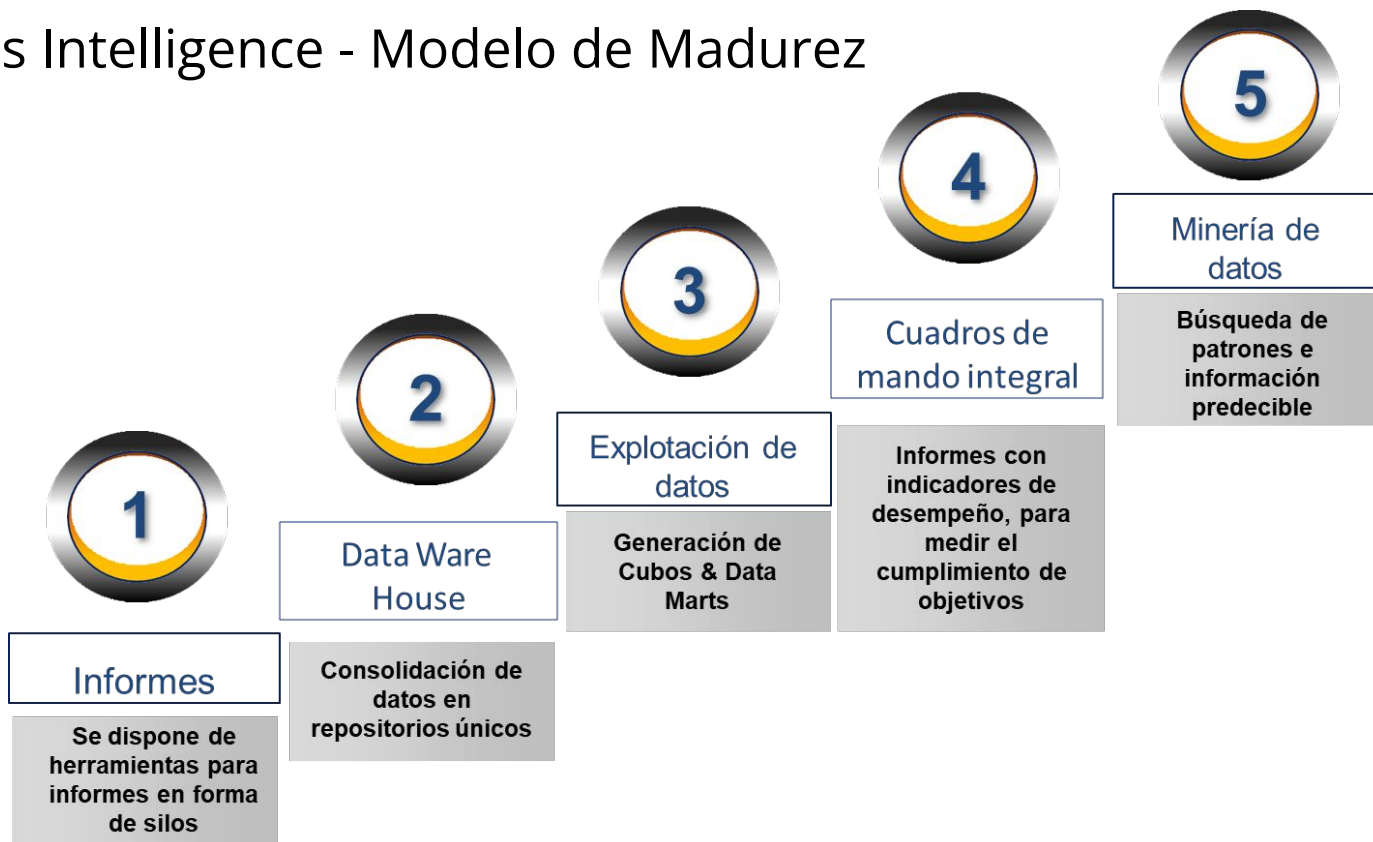
Explotación de Datos

Business Intelligence - Beneficios Esperados



Explotación de Datos

Business Intelligence - Modelo de Madurez



Explotación de Datos

Business Intelligence - Conclusiones

- Un proyecto de BI, no debe ser sólo sobre la exploración de los datos o experimentar con visualizaciones.
- Estas cosas pueden ser útiles en implementaciones de BI maduros, pero es poco probable que mejorar el proceso de negocio se de en las primeras etapas de un proyecto.
- La plataforma de BI, no es solo tecnología sino datos, procesos, personas, que se interrelacionan como un grupo de capacidades focalizadas en el negocio.
- Para que los datos agreguen valor a cualquier empresa, se debe considerar las cuatro dimensiones.

Explotación de Datos

Motivación y aplicabilidad de BI

Dudas / Preguntas



Explotación de Datos

Contenido de la Unidad

1. Introducción a Data Lakes
2. Técnicas de DM y descubrimiento de conocimiento
3. Business Intelligence (BI)
4. Motivación y aplicabilidad de BI
- 5. Curado de datos**
6. Tableros de Control
7. Práctica: Diseño de Dashboards y Tableros
8. Resumen

Explotación de Datos

Curado de Datos

Valor: aumentar el valor de los datos de la organización y las oportunidades de utilizarlos es un business driver.

Riesgo: reducir riesgos y costos asociados por el uso de datos con mala calidad.

Eficiencia y Productividad: se entiende que el ejercicio de la toma de decisiones sobre la base de datos “verdaderos” y completos mejora ambos aspectos.

Explotación de Datos

Curado de Datos

Reputación: la utilización y exposición de datos con mala calidad afecta la imagen de la organización dentro y fuera.

Costos: decisiones incorrectas o mal fundamentadas conllevan aumento de costos innecesarios.

Prevención: el enfoque correcto es sobre prevenir errores, no simplemente en corregir registros erróneos.

Causa Raíz: abordar causas principales en lugar de los síntomas

Explotación de Datos

Curado de Datos - Conceptos Esenciales

La calidad no es un accidente; siempre es el resultado de un esfuerzo de la inteligencia. John Ruskin, escritor británico (1819-1900)

- **Data Quality:** El término DQ se refiere tanto a las características asociadas con los datos de alta calidad como a los procesos utilizados para medir o mejorar la calidad de los datos.
- **Datos de Alta Calidad:** satisfacen las expectativas y necesidades de los consumidores de datos (son aptos para los fines a los que quieren aplicarlos) □ depende del contexto y de las necesidades del consumidor de datos:
 - Desafío 1: no siempre se conocen las expectativas relacionadas con la calidad. Los clientes no pueden articularlos.
 - Desafío 2: Los requisitos cambian con el tiempo.

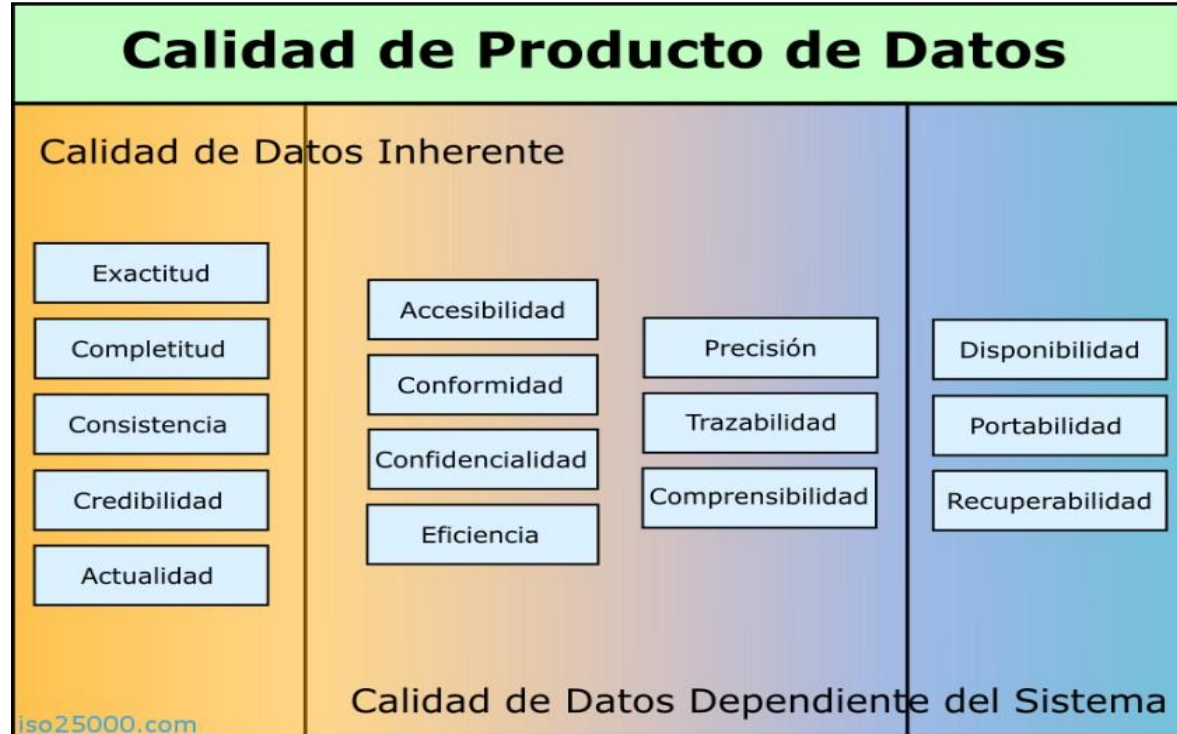
Explotación de Datos

Curado de Datos - Conceptos Esenciales

- **Datos Críticos:** no todos son de igual importancia. Se deben centrar los esfuerzos de mejora en los datos que son más importantes. Impacto directo y medible en las necesidades del negocio: i.e. requerimientos legales o regulatorios, reporte financiero, políticas de negocio, operaciones, estrategia corporativa, diferenciación con la competencia, ... Master Data (por definición)
- **Objetividad:** Los niveles de calidad de los datos deben medirse de forma objetiva y coherente. Las mediciones y la metodología de medición deben compartirse con las partes interesadas, ya que son los árbitros de la calidad (o actores en la mejora).

Explotación de Datos

Curado de Datos - Calidad de Datos



Explotación de Datos

Curado de Datos

Dudas / Preguntas



Explotación de Datos

Contenido de la Unidad

1. Introducción a Data Lakes
2. Técnicas de DM y descubrimiento de conocimiento
3. Business Intelligence (BI)
4. Motivación y aplicabilidad de BI
5. Curado de datos
- 6. Tableros de Control**
7. Práctica: Diseño de Dashboards y Tableros
8. Resumen

Explotación de Datos

Tableros de Control

El **Tablero Business Intelligence, o Tablero de Control Empresarial**, es parte integral de las estrategias de BI que utilizan los negocios de alto nivel. Se trata de soluciones de gestión de información, análisis y visualización de datos fácilmente comprensibles.

Hablamos de herramientas especialmente diseñadas para analizar y visualizar información sobre el comportamiento y transacciones de tu compañía. Además, comprender informes complejos y obtener una visión general del rendimiento.

Un Tablero Business Intelligence es una herramienta para visualizar indicadores importantes sobre el desempeño del negocio. Es decir, reúne una gran cantidad de números en una pantalla gráfica, de manera clara y altamente condensada.

Explotación de Datos

Tableros de Control

LaVoz

Gerencia de Operaciones Area de Producción: Cuadro de Mando Integral

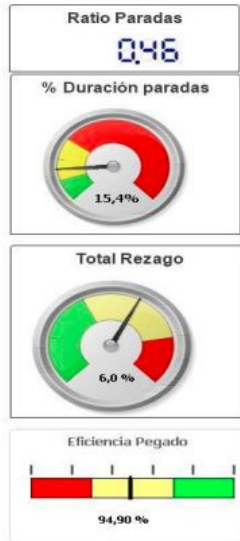
Perspectiva Financiera



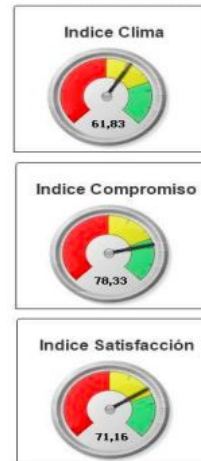
Perspectiva Clientes



Perspectiva Procesos



Perspectiva Aprendizaje y Crecimiento



Explotación de Datos

Tableros de Control - Balanced Scorecard

El **Cuadro de Mando Integral (CMI) o Balanced Scorecard (BSC)**, es una metodología de gerenciamiento presentada en el año 1992 en Harvard Business Review que hace hincapié en una visión más balanceada de las organizaciones, sin tanto foco en la parte financiera.

No deja de ser un Tablero de Control, pero la diferencia es conceptual respecto al alcance de lo que una organización debe **MEDIR** para asegurar su éxito.

Explotación de Datos

Tableros de Control - Balanced Scorecard

Propone analizar las organizaciones en cuatro perspectivas:

- Perspectiva de aprendizaje y crecimiento
- Perspectiva del cliente
- Perspectiva interna
- Perspectiva financiera



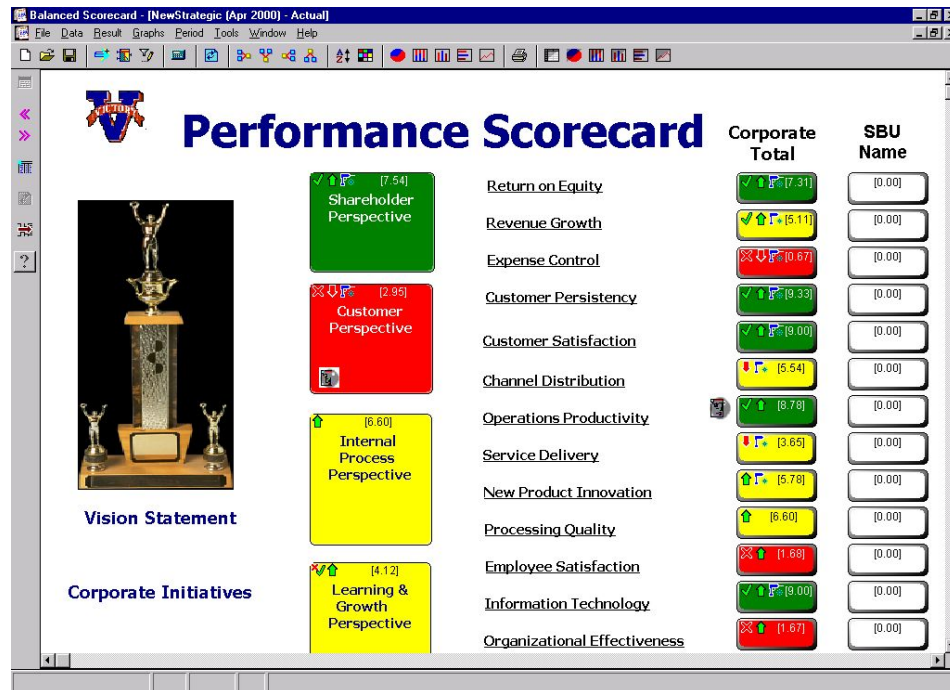
Explotación de Datos

Tableros de Control - Balanced Scorecard

- **Perspectiva financiera:** refleja indicadores de gestión financiera (inversiones, rentabilidad)
- **Perspectiva del cliente:** describe las características del mercado en el cual la empresa compite, y proporciona datos relevantes necesarios para identificar, generar, atraer y fidelizar a los clientes.
- **Perspectiva interna:** mira hacia adentro de la organización e identifica los procesos y actores que generan las iniciativas dentro de ella, sobre todo aquellos críticos y trascendentales en la salud de la empresa.
- **Perspectiva de aprendizaje y crecimiento:** quizá la perspectiva más importante, dado que refleja la proyección a futuro de la organización a través de la infraestructura humana, sistemática y de clima organizacional.

Explotación de Datos

Tableros de Control - Balanced Scorecard



Explotación de Datos

Tableros de Control

Dudas / Preguntas



Explotación de Datos

Contenido de la Unidad

1. Introducción a Data Lakes
2. Técnicas de DM y descubrimiento de conocimiento
3. Business Intelligence (BI)
4. Motivación y aplicabilidad de BI
5. Curado de datos
6. Tableros de Control
- 7. Práctica: Diseño de Dashboards y Tableros**
8. Resumen

Explotación de Datos

Práctica: Diseño de Dashboards y Tableros

Explotación de Datos

Contenido de la Unidad

1. Introducción a Data Lakes
2. Técnicas de DM y descubrimiento de conocimiento
3. Business Intelligence (BI)
4. Motivación y aplicabilidad de BI
5. Curado de datos
6. Tableros de Control
7. Práctica: Diseño de Dashboards y Tableros
8. **Resumen**

Explotación de Datos

Resumen de la Unidad

Durante esta unidad hemos aprendido:

- Introducción a Data Lakes
- Técnicas de DM y descubrimiento de conocimiento
- Business Intelligence (BI)
- Motivación y aplicabilidad de BI
- Curado de datos
- Tableros de Control
- Diseño de Dashboards y Tableros

Explotación de Datos

Bibliografía Recomendada



- Libro 1
- Libro 2
- The Balanced Scorecard: Translating Strategy into Action.
Robert S. Kaplan & David P. Norton, 1996

