

Ingeniería de Datos

Proceso de Desarrollo de las Bases de Datos en la Empresa

Nadia Rodríguez



Agenda

- Terminología
- Desarrollo de las BD dentro del Ciclo de Vida de los Sistemas de Información
- Componentes de un Entorno de Base de Datos
- Ventajas, desventajas riesgos y costos de las BD
- Evolución de las Bases de Datos

Introducción

- Los negocios generan datos...
- **Dato** = recurso organizacional para competir
 - Existe necesidad de información en todos los niveles de la organización
- **Información** = grado de disminución de la incertidumbre



Introducción

- Dato = representación de un mensaje (estructurado o no estructurado)

DEBE SER OBJETIVO

- Información = significado percibido al recibir un mensaje

PUEDE SER SUBJETIVO !

*Ambos términos están estrechamente relacionados, en la práctica ambos términos son frecuentemente intercambiados.

El mundo de los Negocios y los datos

- Problemas:
 - Sobrecarga de información
 - Mucha información genérica
 - Mala calidad de los datos
 - Ausencia de información personalizada y/o relevante para los distintos perfiles que existen en un negocio
 - Falta de retroalimentación oportuna para la mejora de los negocios

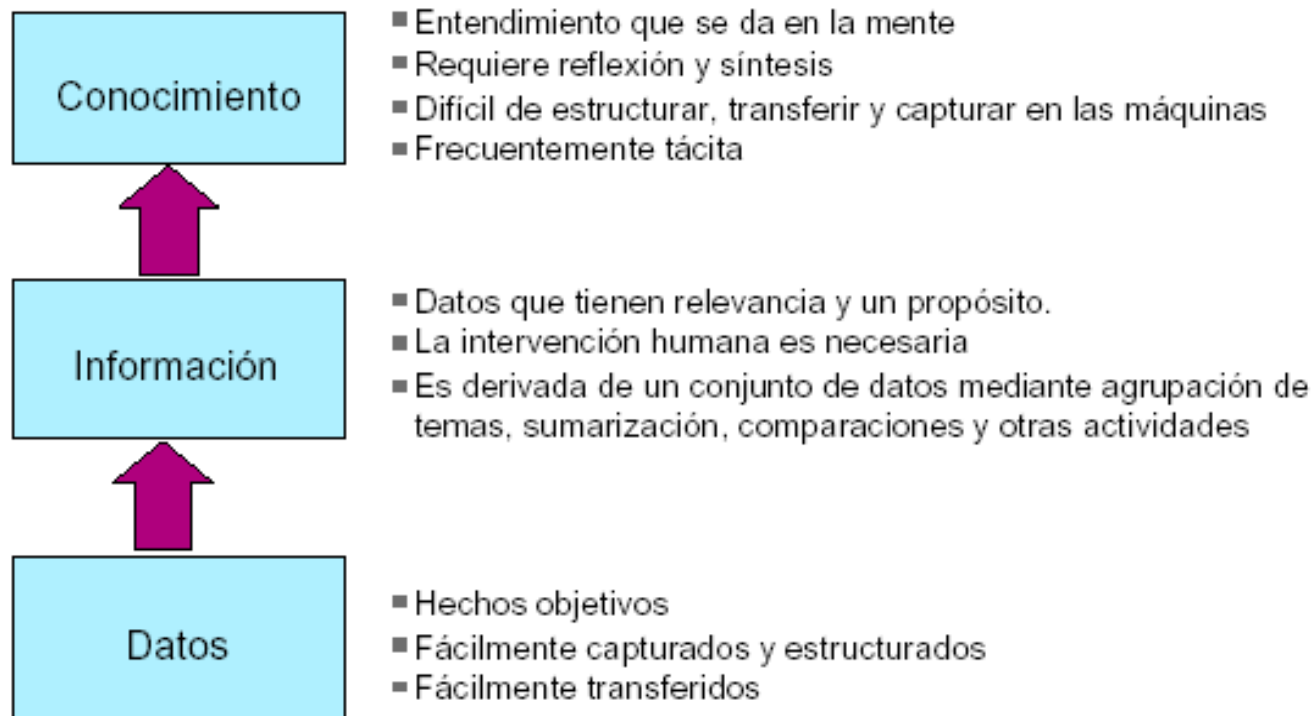


El mundo de los Negocios y los datos

- Nuevos Requerimientos:
 - Los datos son un activo valioso para los negocios
 - Deben ser administrados de manera inteligente



Datos, información y conocimiento



Datos

Dato, en sí mismo tiene poco contenido semántico

Por ejemplo:

Precio:

3.50

Dato, tiene más significado cuando se presenta como parte de una estructura

Por ejemplo:

Código	Nombre	Tipo	Precio	Sabor
001	Cola manzana	Cola	3.50	Manzana

Información

Información, es relevante si hace una diferencia el recibirla o no recibirla

Por ejemplo:

Código	Nombre	Tipo	Precio	Sabor
001	Cola manzana	Cola	3.50	Manzana

Reporte Precios de Colas:
Cola Manzana

2000	2.0
2001	2.7
2002	2.9
2003	3.5

Conocimiento

La información es el origen del conocimiento

Reporte Venta de Colas

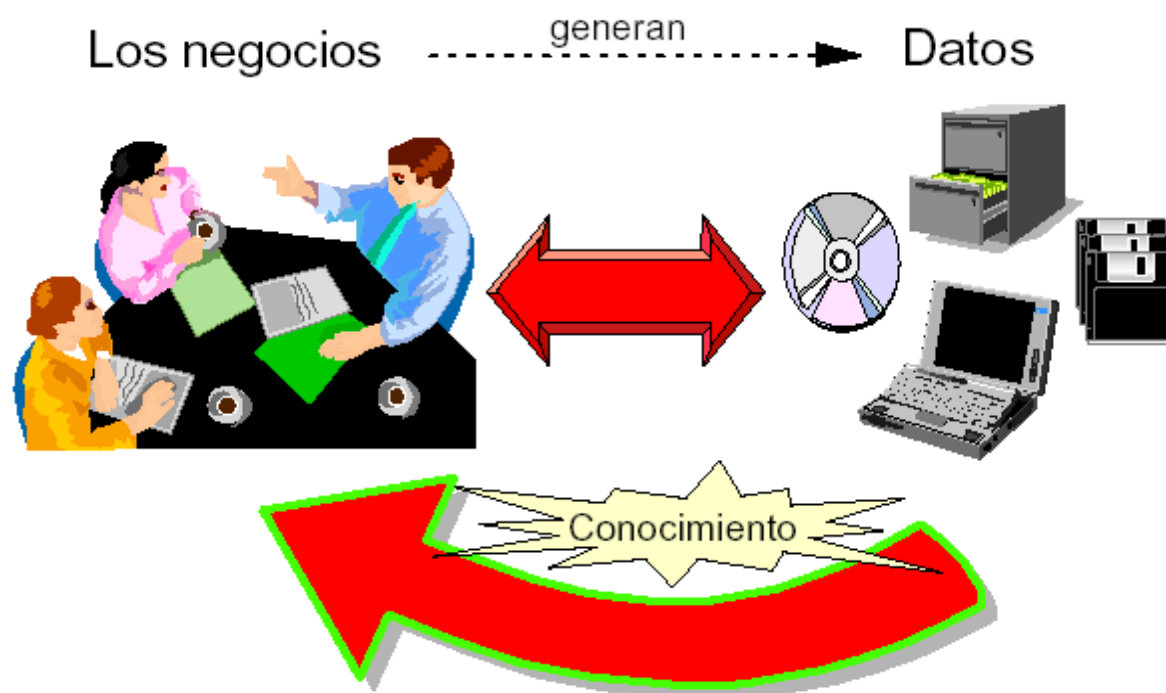
Año	Nombre Cola	Tipo	Precio	Venta
2000	Cola frutilla	Cola	2.5	2'000,000
2001	Cola frutilla	Cola	2.5	1'800,000
2000	Cola manzana	Cola	2.0	2'000,000
2001	Cola manzana	Cola	2.7	2'400,000



Conocimiento

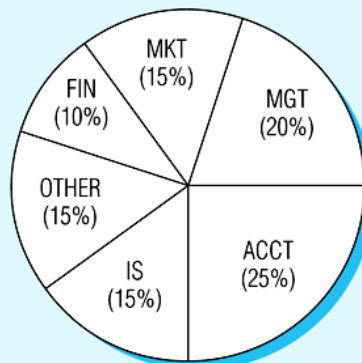
"Si la venta de la cola frutilla ha disminuido a pesar de mantener el mismo precio hay que darle mayor impulso a este sabor"

El resultado



Ejemplos

Class Roster	
Course: MGT 500 Business Policy	Semester: Spring 200X
Section: 2	
Name	ID
Baker, Kenneth D.	32491
Doyle, Joan E.	47619
Finkle, Clive R.	54842
Lewis, John C.	55174
McFerran, Debra R.	40972
Sisneros, Michael	39241



Percent Enrollment by Major (199X)

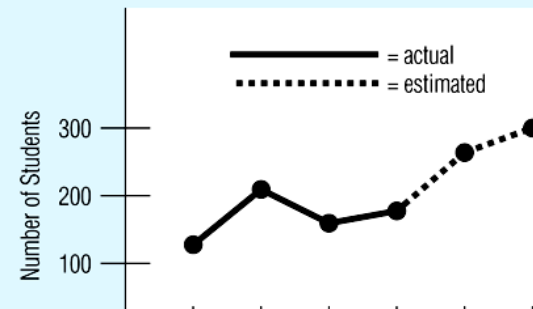


Table 1-1 Example Metadata for Class Roster

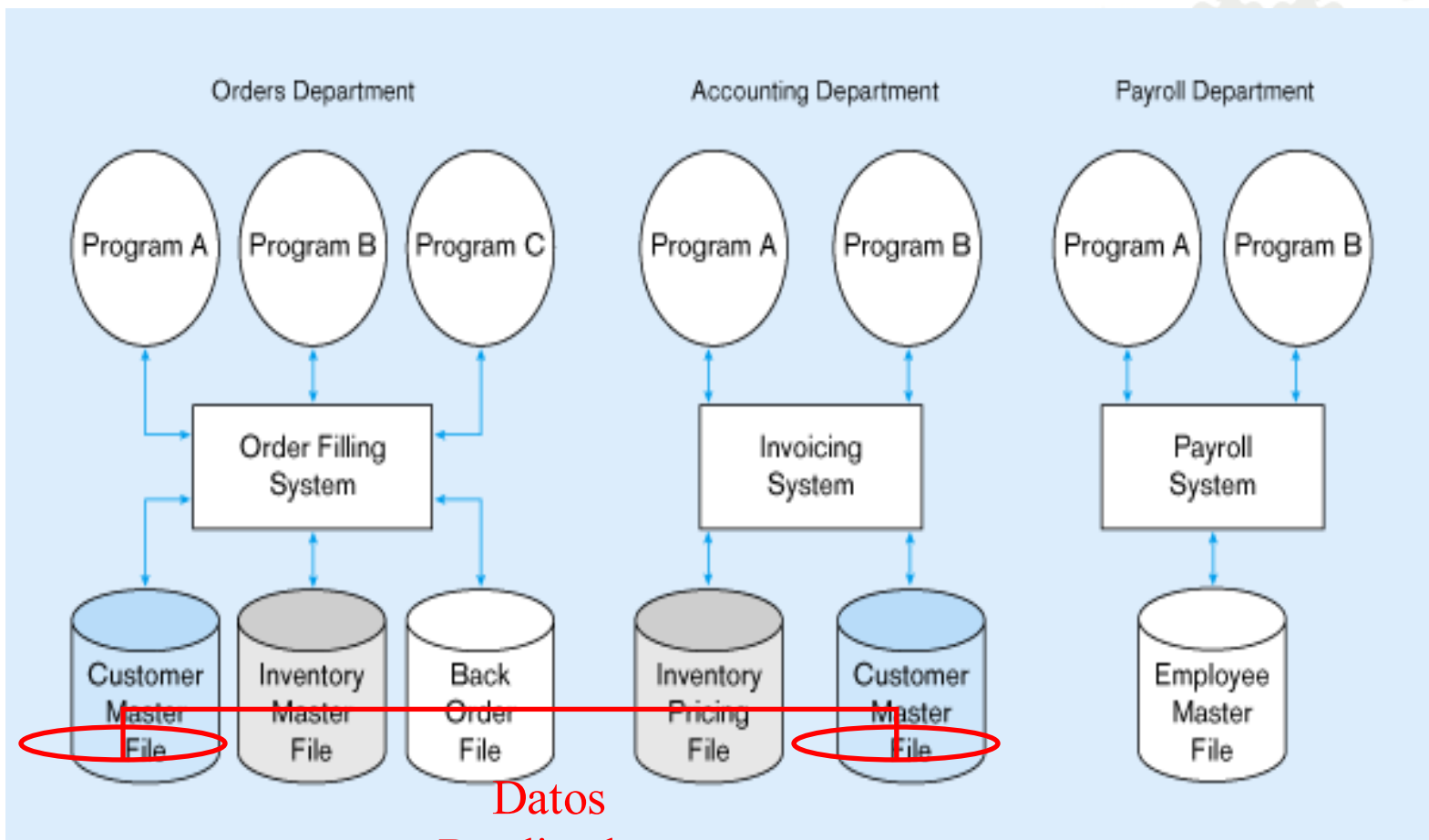
Data Item			Value		
Name	Type	Length	Min	Max	Description
Course	Alphanumeric	30			Course ID and name
Section	Integer	1	1	9	Section number
Semester	Alphanumeric	10			Semester and year
Name	Alphanumeric	30			Student name
ID	Integer	9			Student ID (SSN)
Major	Alphanumeric	4			Student major
GPA	Decimal	3	0.0	4.0	Student grade point average

Metadata: Describe las propiedades y características de los datos y de su contexto.



Sistema Tradicional de Procesamiento de Archivos...

Escenario



Datos
Duplicados

Desventajas del Sistema Tradicional de Procesamiento de Archivos

- Dependencia programa/datos
- Redundancia de datos (Duplicación de datos)
- Limitaciones para compartir datos
- Mayor tiempo de desarrollo de sistemas informáticos y/o programas
- Requiere de mucho mantenimiento de sistemas y/o programas



Desventajas del Sistema Tradicional de Procesamiento de Archivos

Problema con la dependencia programa/datos

- Cada sistema informático y/o programa
 - ✓ Debe ser actualizado con sus propios datos
 - ✓ Necesita incluir código para la metadata de cada archivo
 - ✓ Debe tener sus propias rutinas para la lectura, inserción, actualización y eliminación de datos
- Débil coordinación y control centralizado
- No uso de estándares



Desventajas del Sistema de Procesamiento de Archivos

Problemas con la redundancia de datos

- Espacio desperdiciado por la duplicación de datos
- Los mantenimientos causan muchos dolores de cabeza
- El gran problema:
 - Cuando los datos son cambiados en un archivo, podrían causar inconsistencias
 - Se compromete *la integridad de datos*

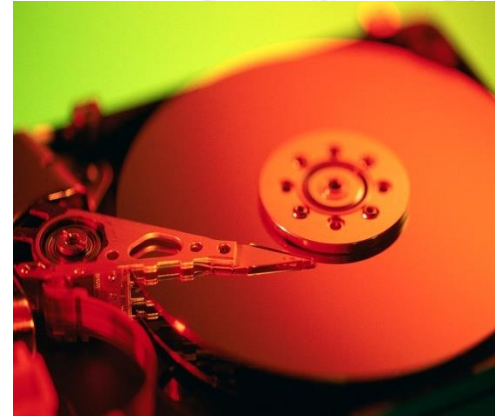


Solución...



Base de Datos

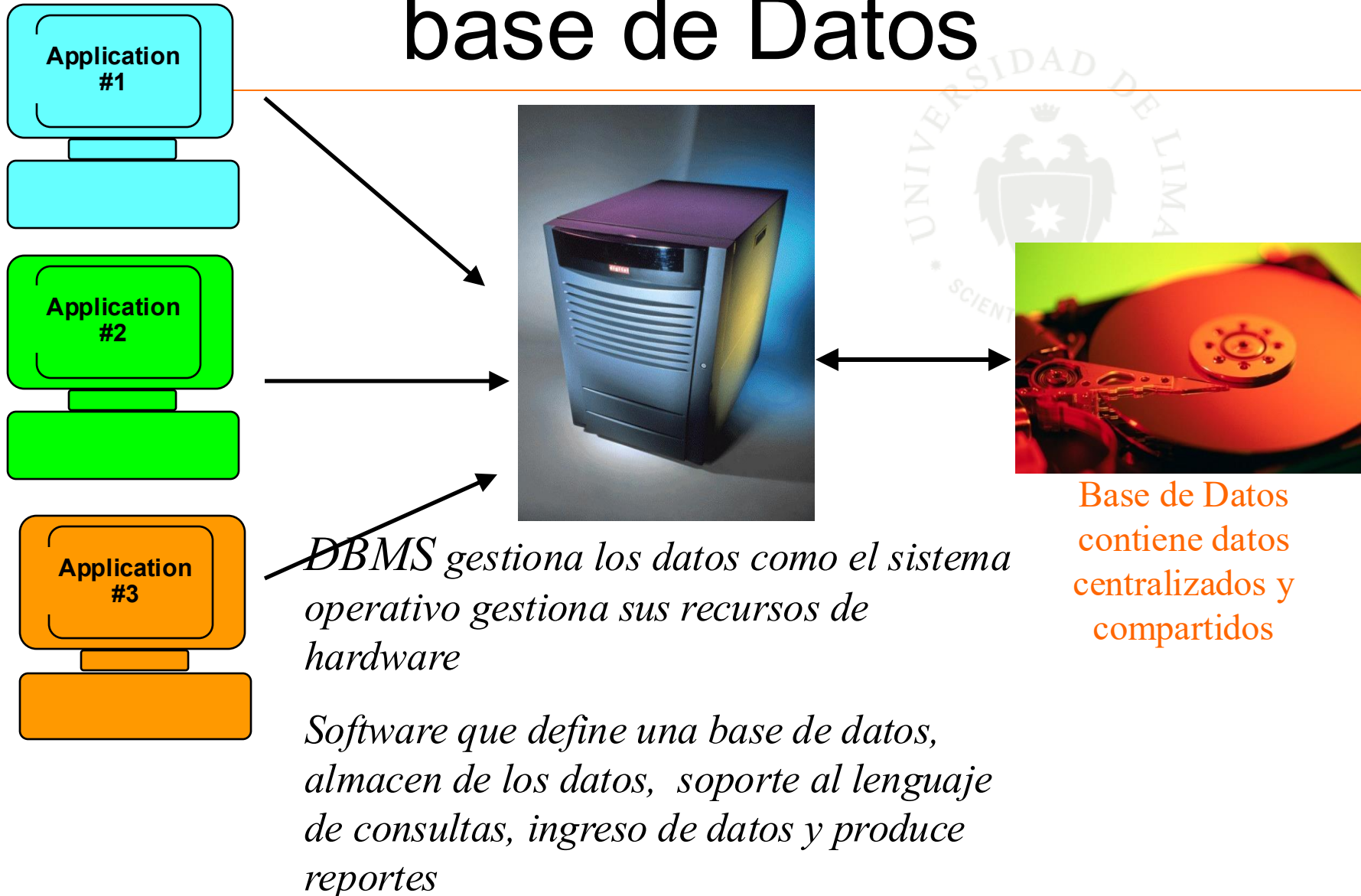
Coleccion de datos almacenados en un formato estandar, diseñado para que pueda ser compartido por multiples usuarios



Requiere de un sistema de gestión de base de datos
Database Management System (DBMS)



Sistema de Gestión de base de Datos



Ventajas de un enfoque orientado hacia Base de Datos

- Independencia programa/datos
 - Los metadatos son almacenados en el DBMS, por lo que las aplicaciones no necesitan preocuparse por los formatos de los datos
 - Las consultas y actualizaciones son gestionadas por el DBMS por lo que los programas no requieren de rutinas de acceso a los datos
 - **Logros:** mejora la productividad en el proceso de desarrollo y mantenimiento de programas y/o sistemas informáticos
- Redundancia mínima de datos
 - Dirigido a mejorar la ***consistencia e integridad***

Ventajas de un enfoque orientado hacia Base de Datos

- Mejora la compartición de datos
 - Diferentes usuarios obtienen diferentes vistas de datos
- Reforzamiento de estándares
 - Todos los accesos a datos son realizados de la misma manera
- Mejoramiento de la calidad de datos
 - Restricciones, reglas de validación de datos
- Mejor accesibilidad/tiempo de respuesta a los datos
 - Uso de un lenguaje estándar de consultas (SQL)
- Seguridad, Backup/Recovery, Concurrencia
 - La recuperación ante desastres es mas sencilla

Costos y riesgos de un enfoque orientado hacia Base de Datos

- Costos iniciales
 - Complejidad y costos de instalación y gestión
 - Costos de conversión
- Costos permanentes
 - Personal especializado
 - Determinación de procesos de backup y recovery
- Conflicto organizacional
 - Viejos hábitos difíciles de vencer




Por que se requiere de un enfoque de Base de Datos?

- Muchas aplicaciones cambian, pero la estructura de los datos es la misma.
- Necesidad de:
 - Requerimientos y consultas *ad hoc* (*a medida*).
 - Reducir los tiempos de desarrollo.
 - Compartir datos.
- Mejorar la calidad de los datos.
- Empowerment de los usuarios.

Construyendo el Sistema adecuado: viabilidad

- | • Costs | Easy to estimate | • Benefits | Hard to value |
|--|------------------|--|---------------|
| <ul style="list-style-type: none">– Up-front/one-time<ul style="list-style-type: none">• Software (\$ millions !)• Hardware• Communications• Data conversion• Studies and Design• Training– On-going costs<ul style="list-style-type: none">• Personnel• Software upgrades• Supplies• Support• Software & Hardware maintenance | | <ul style="list-style-type: none">– Cost Savings<ul style="list-style-type: none">• Software maintenance• Fewer errors• Less data maintenance• Less user training– Increased Value<ul style="list-style-type: none">• Better access to data• Better decisions• Better communication• More timely reports• Faster reaction to change• New products & services– Strategic Advantages<ul style="list-style-type: none">• Lock out competitors | |

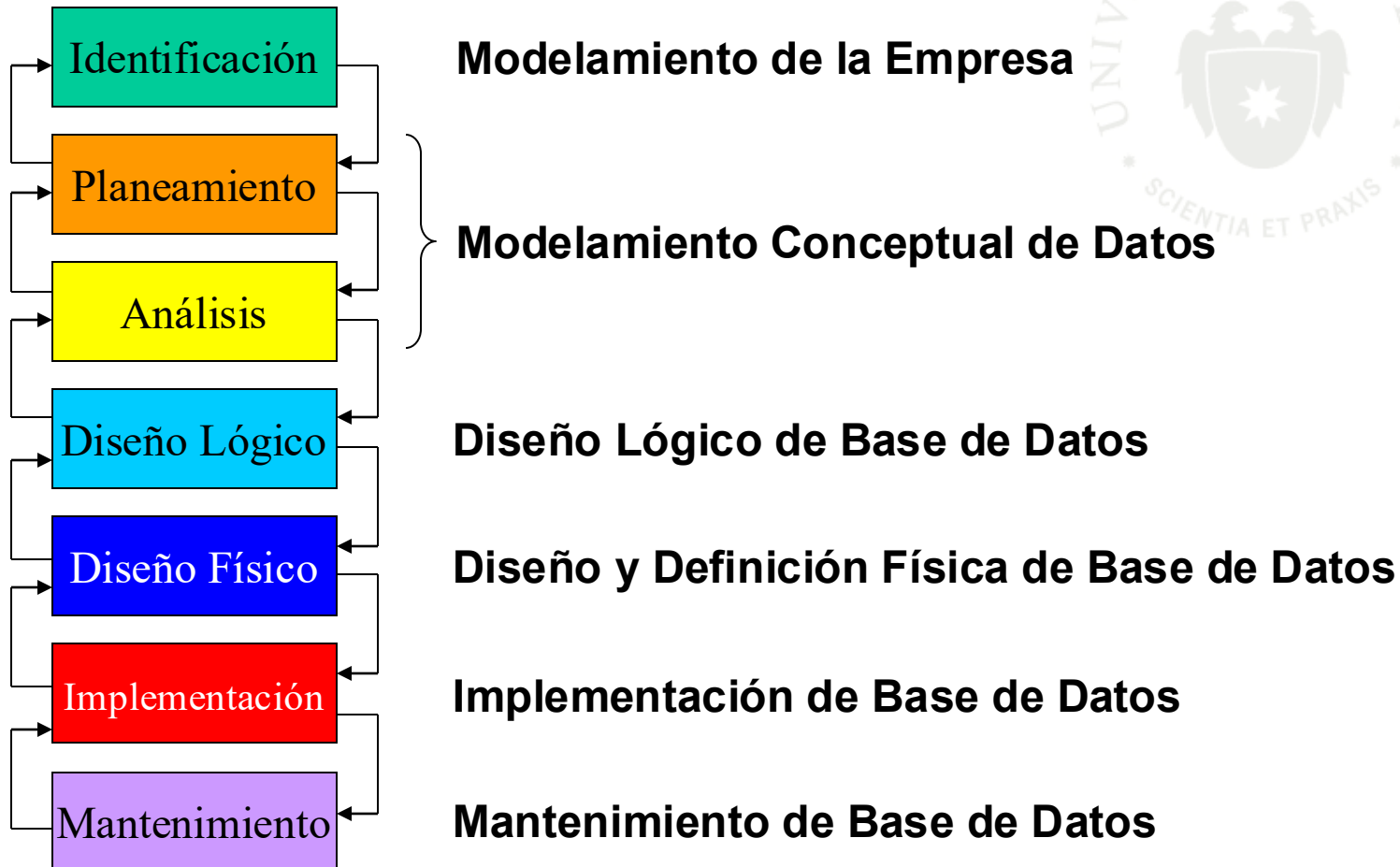


Ciclo de Vida de los Sistemas de Información y las Bases de Datos

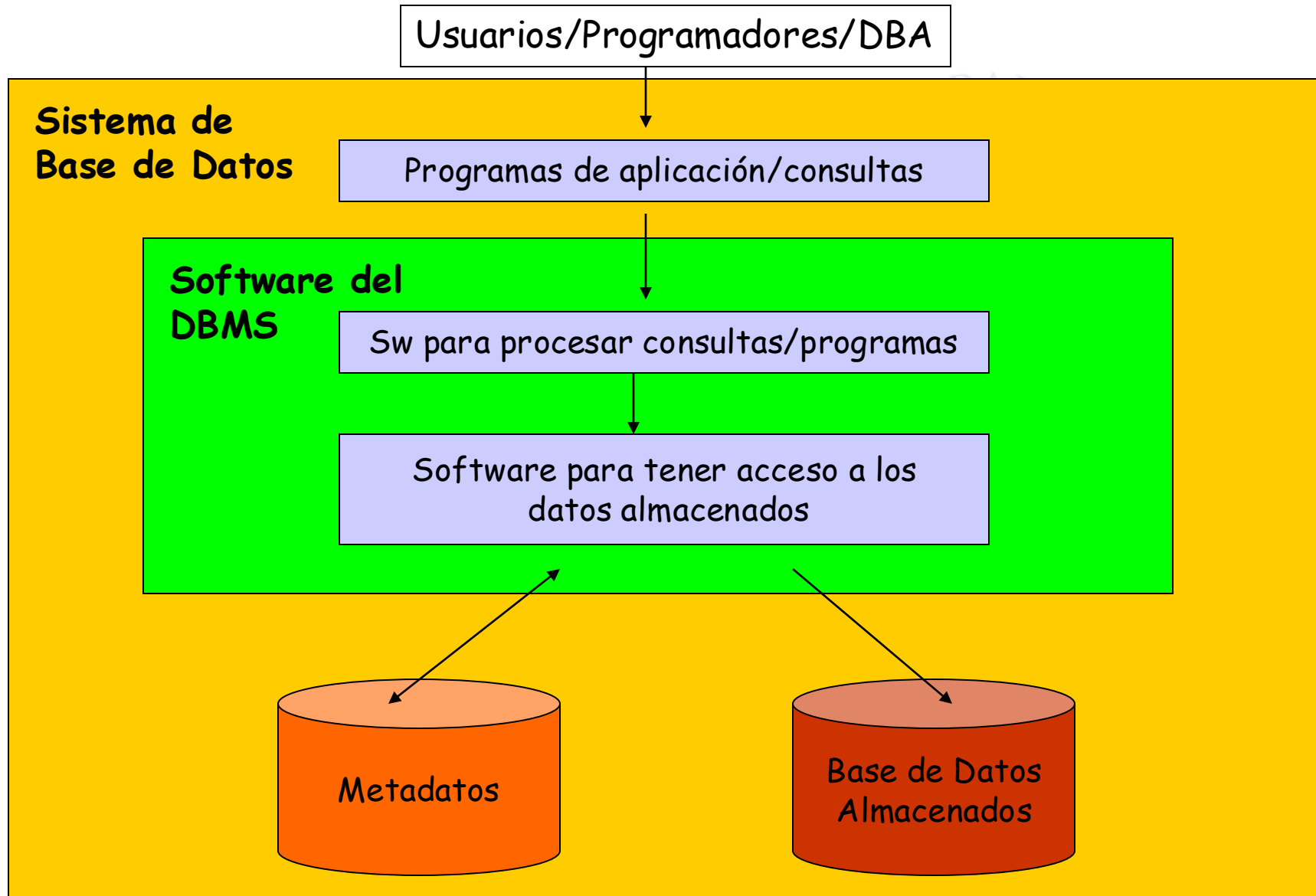
Implementación de Sistemas de Información y Base de Datos

- **System Development Life cycle - SDLC**
 - Proceso de desarrollo detallado, bien planificado.
 - Consume tiempo, pero es comprensivo
 - Ciclo de desarrollo largo
- **Prototipo**
 - Rapid application development (RAD)

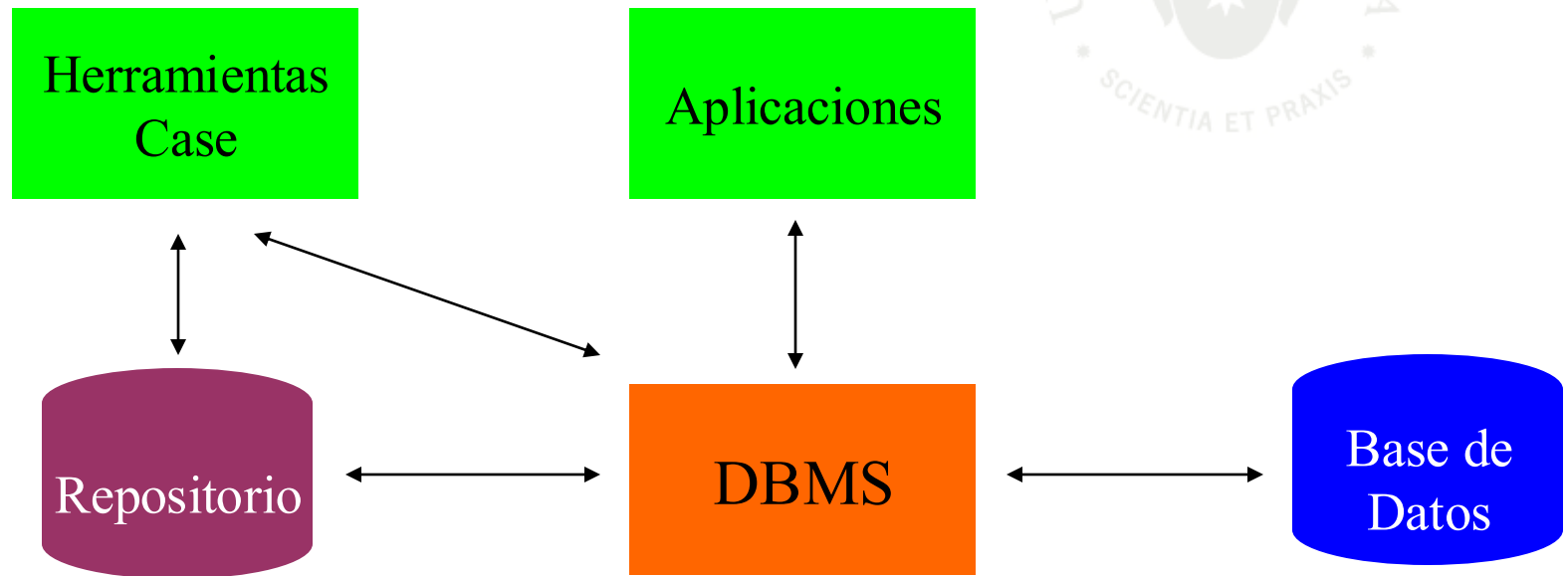
Desarrollo de las Bases de Datos dentro del Ciclo de Vida de los Sistemas de Información (SDLC)



Sistema de Gestión de Base de Datos



Componentes de un Entorno de Base de Datos

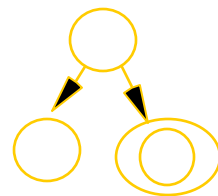
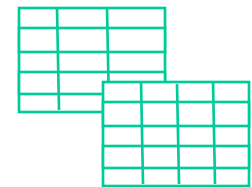
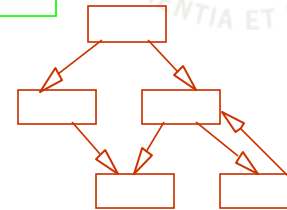
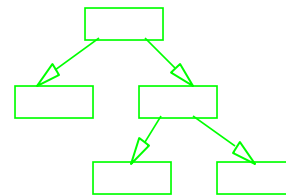


Componentes de un Entorno de Base de Datos

- **Herramienta Case:** herramientas usadas para diseñar BD y aplicaciones.
- **Aplicación:** programa usado para crear, modificar y leer en la BD.
- **Repositorio:** base de conocimientos en donde se guardan definiciones de datos, relaciones, formatos, y otros componentes.
- **Base de Datos:** colección organizada de datos lógicamente relacionada.
- **DBMS:** sistema de software empleado para crear, mantener y proveer información con acceso controlado.

Historia: Base de Datos (BD)

- 1950s Sistemas de ficheros, tarjetas perforadas
- 1960s Jerárquicos
 - IDS, IMS (Sistema Sabre)
- 1970s Redes
 - CODASYL, IDMS
- 1980s Relational
 - INGRES, ORACLE, DB2, Sybase
 - Paradox, dBase
- 1990s Orientado al objeto y objecto relacional
 - O2, GemStone, Ontos
- 199x-2000s Datos en Web, Aplicaciones E-commerce, Tecnologías Wireless, Inteligencia Artificial, XML, NoSQL



Consideraciones para la implementación de una BD

- Complejidad (?)
- Tamaño
- Costo (?)
- Requerimientos adicionales de HW
- Impacto ante fallas (?)
- Recuperación (?)



Terminología

- Transacción
 - Operación completa de lectura o escritura sobre la Base de datos
- Database Administrator (DBA):
 - Administrador de base de datos

