





# Capitulo III

# METODOLOGIA DE DISEÑO DE BASES DE DATOS

Ing. Edgar T. Espinoza R.

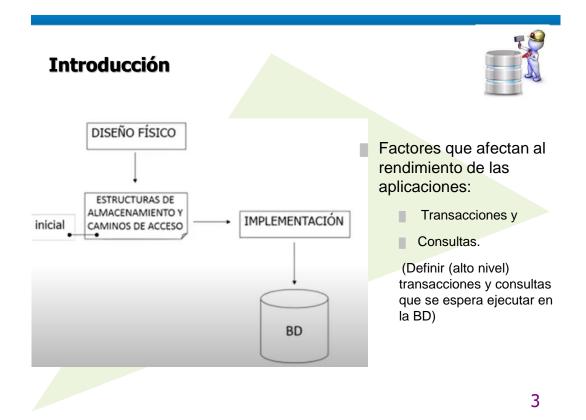
DISEÑO FISICO

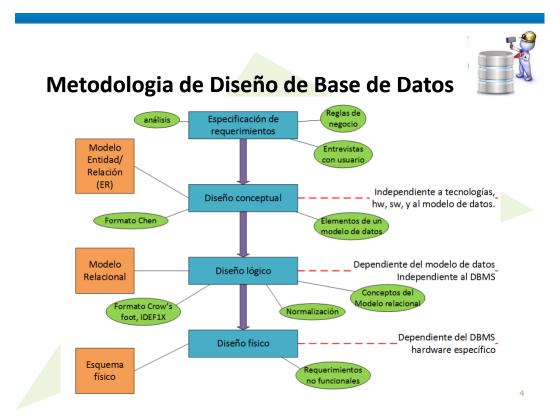




# Introducción

Especificación de estructuras de almacenamiento internas y caminos de acceso específicos para que las diversas aplicaciones que accedan a la BD tengan un buen rendimiento







# **Diseño Conceptual**

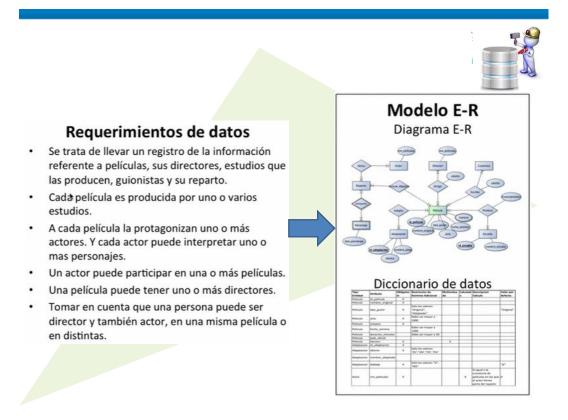
- Proceso de obtener un diseño que sea independiente del modelo de base de datos, del SGBD y de todas las consideraciones físicas.
- Se genera a partir de los requerimientos de datos
- Deriva en un modelo conceptual.
- El más usado es el **Modelo E-R**: basado en la identificación de entidades y sus relaciones.

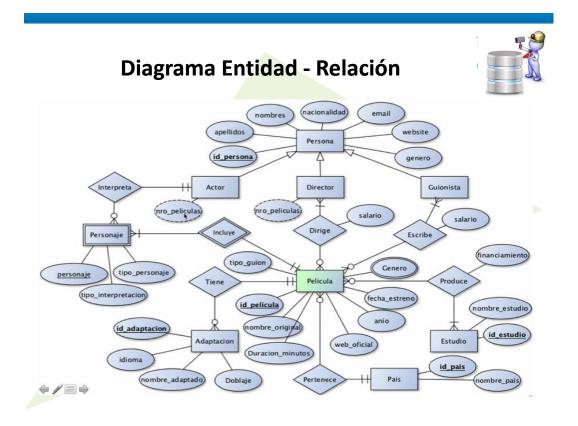
5

# **CASO DE ESTUDIO "FILMES"**



- Dado el siguiente caso de estudio
  - Se trata de llevar un registro de la información referente a películas, sus directores, estudios que las producen, guionistas y su reparto.
  - Cada película es producida por uno o varios estudios.
  - A cada película la protagonizan uno o más actores. Y cada actor puede interpretar uno o mas personajes.
  - Un actor puede participar en una o más películas.
  - Una película puede tener uno o más directores.
  - Tomar en cuenta que una persona puede ser director y también actor, en una misma película o en distintas.
  - Referencias:





## Diccionario de datos

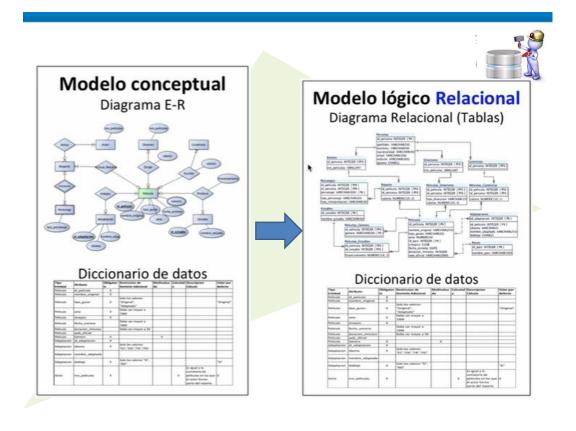


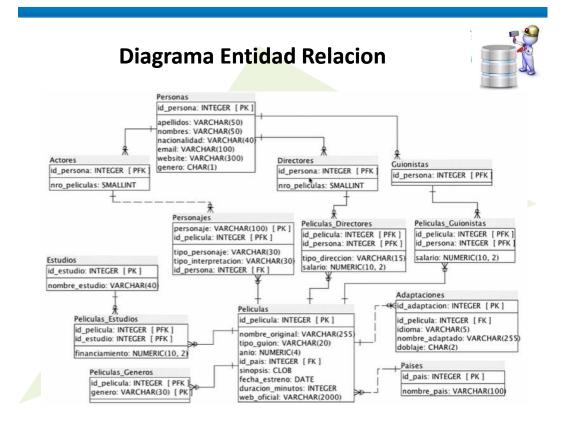
Tipo Atributo		Obligat	Restriccion de Dominio Adicional	Multiva	Calcul	Descripcion Cálculo	Valor por	
Entidad	orio		Restriccion de Dominio Adicional	luado	ado	Descripcion Calculo	defecto	
Pelicula	nombre_original	X.						
Pelicula	tipo_guion	х	Solo los valores: "Original", "Adaptado"				"Original"	
Pelicula	anio	Х	Debe ser mayor o igual a 1900					
Pelicula	sinopsis	Х						
Pelicula	fecha_estreno		Debe ser mayor o igual a 1900					
Pelicula	duracion_minutos		Debe ser mayor a 30					
Pelicula	genero	Х		Х				
Adaptacion	idioma	х	Solo los valores: "ES","EN","FR","PG"					
Adaptacion	doblaje	Х	Solo los valores "SI", "NO"				"SI"	
Actor	nro_peliculas	х			х	Es igual a la sumatoría de películas en las que el actor forma parte del reparto	0	
Director	nro_peliculas	х			х	Es igual a la sumatoría de películas en las que el actor forma parte del reparto	0	
Persona	apellidos	Х						
Persona	nombres	Х						
Persona	genero		Solo valores "M", "F"					
Personaje	tipo_interpretacion		Solo los valores: "Actuación","Voz"					
Estudio	nombre_estudio	Х						
Pais	nombre_pais	Х						
Dirige	salario		No puede ser negativo					
Escribe	salario		No puede ser negativo					
Produce	financiamiento		No puede ser negativo					

# Diseño Logico



- Proceso de obtener un diseño apegado a un modelo de base de datos específico (Relacional, Orientado a Objetos, Jerárquico, Orientado a Grafos, etc.)
- Pero aun es independiente de un SGBD específico y de todas las consideraciones físicas de implementación.
- · Se genera a partir del modelo conceptual





#### Diccionario de Datos



Tabla	Campo	Obligat orio	Restriccion de Dominio Adicional	Calcul	Descripcion Cálculo	Valor por defecto
Peliculas	nombre_original	X				
Peliculas	tipo_guion	х	Solo los valores: "Original", "Adaptado"			"Original"
Peliculas	anio	X	Debe ser mayor o igual a 1900			
Peliculas	sinopsis	Х				
Peliculas	fecha_estreno		Debe ser mayor o igual a 1900			
Peliculas	duracion_minutos		Debe ser mayor a 30			
Adaptaciones	idioma	х	Solo los valores: "ES","EN","FR","PG"			
Adaptaciones	id_pelicula	Х				
Adaptaciones	doblaje	X	Solo los valores "SI", "NO"			"SI"
Actores	nro_peliculas	х		x	Es igual a la sumatoría de películas en las que el actor forma parte del reparto	0
Directores	nro_peliculas	х		x	Es igual a la sumatoría de películas en las que el actor forma parte del reparto	0
Personas	apellidos	X				
Personas	nombres	X				
Personas	genero		Solo valores "M", "F"			
Personajes	tipo_interpretacion		Solo los valores: "Actuación","Voz"			
Personajes	id_persona	X				
Estudios	nombre_estudio	X				
Paises	nombre_pais	X				
Peliculas_Directores	salario		No puede ser negativo			
Peliculas_Guionistas	salario		No puede ser negativo			
Peliculas_Estudios	financiamiento		No puede ser negativo			

# Diseño Físico



- Mientras el diseño lógico se preocupa del "Que" el diseño físico ocupa del "Como"
- Describe todas las especificaciones de implementación de la base de datos en un entorno específico
- Implica definir y conocer el SGBD a usar y su entorno de implementación: hardware y sistema operativo
- · Aparece el rol DBA

#### Diseño Físico

#### Proceso

- A. Traducir Modelo Lógico a DDL (Lenguaje de Definición de Datos)
- B. Diseñar organización física de los datos
  - Analizar transacciones
  - Establecer organización física de archivos
  - Identificar de índices requeridos
  - Calcular espacio en disco
- C. Definir vistas de usuario
- D. Diseñar mecanismos de seguridad
- E. Optimizar rendimiento para acceso a datos
- F. Monitorizar la base de datos

15

# 

# Traducir Modelo Logico a DDL



#### DDL (Oracle)

CREATE TABLE adaptaciones (
 id\_adaptacion INTEGER NOT NULL,
 id\_pelicula INTEGER NOT NULL,
 idioma VARCHAR(5) NOT NULL,
 nombre\_adaptado VARCHAR(255),
 doblaje CHAR(2),
 CONSTRAINT adaptaciones\_pk
 PRIMARY KEY (id\_adaptacion)

ALTER TABLE adaptaciones ADD
CONSTRAINT Peliculas\_Adaptaciones\_fk
FOREIGN KEY (id\_pelicula)
REFERENCES Peliculas (id\_pelicula)
ON DELETE NO ACTION
ON UPDATE NO ACTION;

16

# Diseñar la organización física Analizar Transacciones



				Matriz (	ruzada	transac		relacion	es					
	TABLAS													
Transacción	Persona	Actores	Directo	Guionis	Pelicula s	Persona ies	Mark to the latest latest	Película s_Guio nistas	Estudio	The second second	Pelicula s_Gene ros	AND DESCRIPTION OF THE PARTY OF	Paises	# Tablas por transacción
Registrar una nueva película	х	х	x	х	х	х	х	х	х	х	х	х	x	13
Gestionar Actores	х	x												2
Gestionar Directores	х		x											2
Gestionar Guionistas	x			х										2
Actualizar Informacion de Películas	х	х	х	х	x	х	х	х	х	х	х	х	х	13
Mostrar ficha informativa de una película	х	x	x	х	х	х	х	х	х	х	х	х	х	13
Mostrar Personajes Interpretados por un actor	x	х			x	х								4
Numero de películas dirgidas por un director	х		x											2
Listar peliculas adaptadas a un idioma (ej ES)					x							x		2
Listar Películas según el género					х						х			2
Listar Películas con mayor presupuesto					х					х				2
Listar actores mejor pagados	х	х			х									3
# Transacciones x tabla	9	6	5	4	8	4	3	3	3	4	4	4	3	

# Diseñar la organización física



# **Organizar Archivos**

- Elegir una organización adecuada de archivos
- Cada SGBD tiene su propia estructura física de almacenamiento
- Por ejemplo, en Oracle:
  - Tablespaces
  - Data files
  - Redolog files
  - Archive log files

# Diseñar la organización física Identificar Indices Requeridos



- Los índices son estructuras ordenadas que permiten que las consultas sean más eficientes
- Se evalúan para campos que son frecuentemente usados para filtrar, combinar, ordenar o agrupar información.
- Llaves primarias (PK) siempre deben tener índice asociado
- Ocupan espacio en disco

19

El uso de índices se emplea para implementar algunas restricciones que ofrece el modelo relacional.



- Un índice es una estructura de datos. Requiere almacenamiento al igual que una tabla.
- El principal objetivo de un índice es el incrementar el desempeño de la base de datos principalmente para realizar la extracción o explotación de los datos en el menor tiempo posible.

Un índice se emplea principalmente para:

- Localizar datos sin tener que recorrer todo el contenido de una tabla. Para realizar el ordenamiento de datos de uno o más campos
- A través del ordenamiento, para garantizar la no duplicidad de los valores de un campo empleados en especial en la PK.

Ing. Edgar Espinoza R.

#### Elementos de un Índice



El concepto de un índice es muy similar al de un libro, esta formado por 2 elementos:

- Una llave (key)
- Un valor.

Índice en una Base de datos:

- Llave → El valor del dato que se quiere buscar. Por ejemplo, obtener todos los registros con email = juan@m.com.
- Valor → Un identificador único que indica la ubicación en disco donde se encuentra el registro solicitado. En Oracle se le conoce como row\_id.

21

#### Row\_ids en Oracle.



- Oracle utiliza una codificación en base 64 para representar las direcciones físicas de cada registro haciendo uso de una cadena de 18 caracteres.
- El DBMS agrega una columna a cada tabla llamada ROW\_ID en la que se almacena esta cadena.
- Los ROW ID están compuestos de cuatro partes:
  - OOOOOO: Data Object Number o ID que identifica el segmento de la base de datos.
  - FFF: El número de data file que contiene los datos del registro relativo a un tablespace.
  - BBBBBB: El bloque de datos que contiene al registro. Los números de bloque son relativos al data file al que pertenecen, no al tablespace. Dos registros con el mismo número de bloque pueden residir en dos data files diferentes del mismo tablespace.
  - RRR: El número de registro en el bloque.

Un ROW\_ID representa la estrategia más eficiente para localizar un registro en la base de datos.

La siguiente tabla muestra los datos de los clientes de una empresa.



#### **CLIENTES**

CLIENTE_ID	NOMBRE	AP_PATERNO	AP_MATERNO	ENTIDAD	SEXO	CODIGO_POSTAL	EMAIL
1000	JUAN	LUNA	JUAREZ	ZAC	Μ	32948	juan@m.com
1002	MARIO	MARTINEZ	AGUILAR	AGS	M	00293	mario@h.com
1003	ALEJANDRO	PEREZ	MORALES	GRO	М	02934	ale@hd.com
1004	HUGO	LINARES	HURTADO	DF	Μ	20398	hg@aol.com
1005	TANIA	SANCHEZ	GUTIERREZ	GRO	F	98909	ts@em.com
1006	ALONSO	LUGO	OLVERA	AGS	M	02934	alo@yo.com
1007	MIRIAM	MONDRAGON	RUBIO	YUC	F	02934	lug@tr.com
1008	JULIETA	ZAVALA	YAÑEZ	CHIH	F	02342	miri@msn.com
1009	JUAN	BENITEZ	LOPEZ	NL	Μ	02349	beno@su.com
1010	MARIA	JIMENEZ	BALDERRAMA	GRO	F	03453	mja@aol.com

#### ROWIDs de la tabla CLIENTE

ROW_ID	CLIENTE_ID	NOMBRE	AP_PATERNO	AP_MATERNO	ENTIDAD	SEXO	CODIGO_POSTAL	EMAIL
AAASSdAAEAAAASdAAA	1000	JUAN	LUNA	JUAREZ	ZAC	M	32948	juan@m.com
AAASSdAAEAAAASdAAB	1002	MARIO	MARTINEZ	AGUILAR	AGS	M	00293	mario@h.com
AAASSdAAEAAAASdAAC	1003	ALEJANDRO	PEREZ	MORALES	GRO	М	02934	ale@hd.com
AAASSdAAEAAAASdAAD	1004	HUGO	LINARES	HURTADO	DF	М	20398	hg@aol.com
AAASSdAAEAAAASdAAE	1005	TANIA	SANCHEZ	GUTIERREZ	GRO	F	98909	ts@em.com
AAASSdAAEAAAASdAAF	1006	ALONSO	LUGO	OLVERA	AGS	М	02934	alo@yo.com
AAASSdAAEAAAASdAAG	1007	MIRIAM	MONDRAGON	RUBIO	YUC	F	02934	lug@tr.com
AAASSdAAEAAAASdAAH	1008	JULIETA	ZAVALA	YAÑEZ	CHIH	F	02342	miri@msn.com
AAASSdAAEAAAASdAAI	1009	JUAN	BENITEZ	LOPEZ	NL	М	02349	beno@su.com
AAASSdAAEAAAASdAAJ	1010	MARIA	JIMENEZ	BALDERRAMA	GRO	F	03453	mja@aol.com VV

# **Identificar Índices Requeridos**



#### Ejemplo

#### Listar las películas según el género

#### Consulta SQL:

SELECT p.nombre\_original, p.anio, p.fecha\_estreno
FROM peliculas p INNER JOIN peliculas\_generos g
ON p.id\_pelicula = g.id\_pelicula
WHERE g.genero = "DRAMA"
ORDER BY p.nombre\_original;



CREATE INDEX nombre\_original\_idx ON peliculas
(nombre\_original);

24

# Diseñar la organización física Calcular Espacio en Disco



	ESPACIO TABLAS PROYECTADO A 5 AÑOS									
	Tamaño	Care	ga Inicial	Carga Diaria	Carga Diaria Proyeccion 5 años			Total a 5 años		
Relaciones	aproximado de cada registro (en Bytes)	Registros	Bytes	Registros	Registros	Bytes	Bytes	МВ		
Personas	555	21,000	11,655,000	2.1	3,780	2,097,900	13,752,900	13.12		
Actores	16	15,000	240,000	1.5	2,700	43,200	283,200	0.27		
Directores	16	3,000	48,000	0.3	540	8,640	56,640	0.05		
Guionistas	14	3,000	42,000	0.3	540	7,560	49,560	0.05		
Peliculas	6,510	250,000	1,627,500,000	30.0	54,000	351,540,000	1,979,040,000	1,887.36		
Personajes	178	3,750,000	667,500,000	450.0	810,000	144,180,000	811,680,000	774.08		
Peliculas_Directores	41	1,000,000	41,000,000	120.0	216,000	8,856,000	49,856,000	47.55		
Películas_Guionistas	26	375,000	9,750,000	45.0	81,000	2,106,000	11,856,000	11.31		
Estudios	54	500	27,000	0.0	54	2,916	29,916	0.03		
Peliculas_Estudios	26	500,000	13,000,000	60.0	108,000	2,808,000	15,808,000	15.08		
Peliculas_Generos	44	500,000	22,000,000	60.0	108,000	4,752,000	26,752,000	25.51		
Adaptaciones	280	1,250,000	350,000,000	150.0	270,000	75,600,000	425,600,000	405.88		
Paises	114	200	22,800	0.0	0	Activa <sub>0</sub>	22,800	0.02		
TOTAL			2,742,784,800			592,002,216	3,334,787,016	<b>3,180.30</b>		

# Diseñar la organización física



# Calcular Espacio en Disco

TOTALES PROYECTADO A 5 AÑOS (en Mbytes)						
Espacio tablas	3,180.30					
Espacio índices	539.55					
Subtotal	3,719.85					
Espacio adicional (20%)	743.97					
TOTAL	4,463.82					

26

#### **Definir Vistas de Usuario**



- Una vista es una proyección limitada de los datos, principalmente para fines de consulta.
- · Su objetivo:
  - Independencia de los datos
  - Restringir el acceso solo a ciertos datos y operaciones
  - Reducir complejidad de las consultas
- En la práctica las vistas suelen usarse más para discriminar aplicaciones que usuarios.

27

# Diseñar Mecanismos de Seguridad



- Es responsabilidad del DBA, implementar todas las medidas para garantizar:
  - la disponibilidad de la base de datos (en tiempo y en rendimiento)
  - la integridad de los datos
  - la integridad de los equipos
  - la no alteración indebida de los datos
  - la no fuga de información

## Diseñar Mecanismos de Seguridad



### **Ejemplo**

- Implica aplicar técnicas de desnormalización para introducir redundancia en la base de datos
- Busca mejorar tiempos de respuesta y complejidad en operaciones de consulta
- Aumenta riesgo de inconsistencias en los datos

29

# **Optimizar Rendimiento**



- Implementar servidor de respaldo.
- Contar con redundancia eléctrica
- Implementar arreglos de discos en servidores (RAID)
- Acceso físico a servidor únicamente al DBA
- Usar de vistas para consulta de datos
- Encriptar campos confidenciales
- Acceso a base de datos solo desde la intranet



#### Monitorizar la Base de Datos

- Realizar un seguimiento de la operación de la BD
- Depende del SGBD: en la mayoría es posible activar utilidades de auditoría o instalar herramientas de monitoreo.
- Se monitorea: memoria, procesador y disco.
- Si se detectan problemas de rendimiento y ralentización de operaciones:
  - 1. Optimizar consultas
  - Revisar modelo de datos
  - 3. Revisar configuración física de la BD (aumentar memoria, reorganizar y desfragmentar archivos)
  - 4. Potenciar servidor
  - 5. Introducir redundancia de datos (desnormalizar)

31







Ing. Edgar Espinoza R.