

PROYECTO FINAL

BDD empleada para administrar el negocio de la empresa NetMAX – PARTE 1

Notas antes de iniciar:

- El proyecto se deberá realizar con **2 nodos físicos reales** lo que genera un total de 4 PDBs.
- Puede realizarse de manera individual con 2 opciones posibles:
 - Emplear 2 equipos
 - Emplear 1 equipo con instalación nativa (2 PDBs nativas) y una instalación con máquina virtual (2 PDBs en la máquina virtual).
- Para no depender de una máquina remota, se pueden crear 2 PDBs adicionales empleadas únicamente para efectos de prueba. El único cambio para hacer “switch” entre las 2 PDBs de prueba y los 2 reales es la modificación de las definiciones de las ligas. Las PDBs adicionales se pueden crear empleando DBCA.
- Finalmente, el proyecto se estará realizando por partes. Se deberán realizar las actividades solicitadas en cada documento.
- Al momento de realizar la entrega del proyecto se deberá entregar la siguiente documentación:
 - Modelo relacional global
 - Modelo relacional local de cada sitio
 - Tabla de asignación de sitios (se explica más adelante)
 - Tabla de fragmentación (se explica más adelante)

Nota importante:

- Durante la entrega del proyecto se hará una demostración de la BDD que incluye la ejecución de todos los scripts desde la creación del usuario. Los integrantes deberán explicar el contenido de cada script SQL. ***El integrante que explique cada archivo será seleccionado al azar.*** De no contestar adecuadamente se tendrán puntos menos en el proyecto. Si se acumulan 3 o más preguntas sin contestar, ***se anula el proyecto.***

1.1. DEFINICIÓN DEL CASO DE ESTUDIO.

Considere las siguientes reglas de operación de una empresa global que ofrece servicios de Video Streaming a través de Internet llamada NetMax.

1.1.1. Requerimientos funcionales.

NetMax cuenta con un catálogo de programas de TV que ofrece a sus clientes clasificados en 3 grandes grupos: Documentales, Películas y Series. Para cada programa independiente de su tipo, se almacenan los siguientes datos: folio alfanumérico único de 13 caracteres, nombre hasta de 100 caracteres, descripción de hasta 500 caracteres. Los programas pasan por ciertos estados dependiendo su ciclo de vida útil:

- PROGRAMADO: Los datos del programa han sido registrados, pero aún no está disponible al público.
- DISPONIBLE: El programa está disponible para que los clientes puedan consumirlo a través de algún cliente de Video Streaming.
- SATURADO: El sistema ha detectado una gran demanda de clientes que quieren acceder a dicho programa, por lo que se actualiza su estado a SATURADO.
- NO DISPONIBLE: El programa aún está registrado en la BD, pero ya no está disponible para que nuevos clientes lo puedan consumir. Solo los clientes existentes pueden continuar consumiéndolo.
- OBSOLETO: el Programa está registrado en la BD, pero ningún cliente lo está consumiendo actualmente, por lo que sus archivos pueden ser eliminados de la BD sin problemas.

La empresa desea almacenar el estado actual y la fecha de cambio de estado para cada programa, así como la historia de cambios de estado a través del tiempo.

Para los documentales se almacena la temática, la duración en minutos, y un pequeño video de 3 minutos (tráiler) que explica su contenido. Cada documental se asocia a un país en el cual fue desarrollado. La empresa cuenta con un catálogo en el que se cuenta con clave del país, nombre y continente (AME, EUR, ASIA, AFR, OCE).

Para las películas se almacena la duración, sinopsis, clasificación (A,B,C,D). No se cuenta con un catálogo de clasificaciones, pero si se requiere que la BD verifique que solo se especifique alguno de estos 4 valores. Para poder entender algunas películas, se le recomienda al usuario ver otra película llamada “película antecedente”. Se requiere asociar esta película antecedente en la BD.

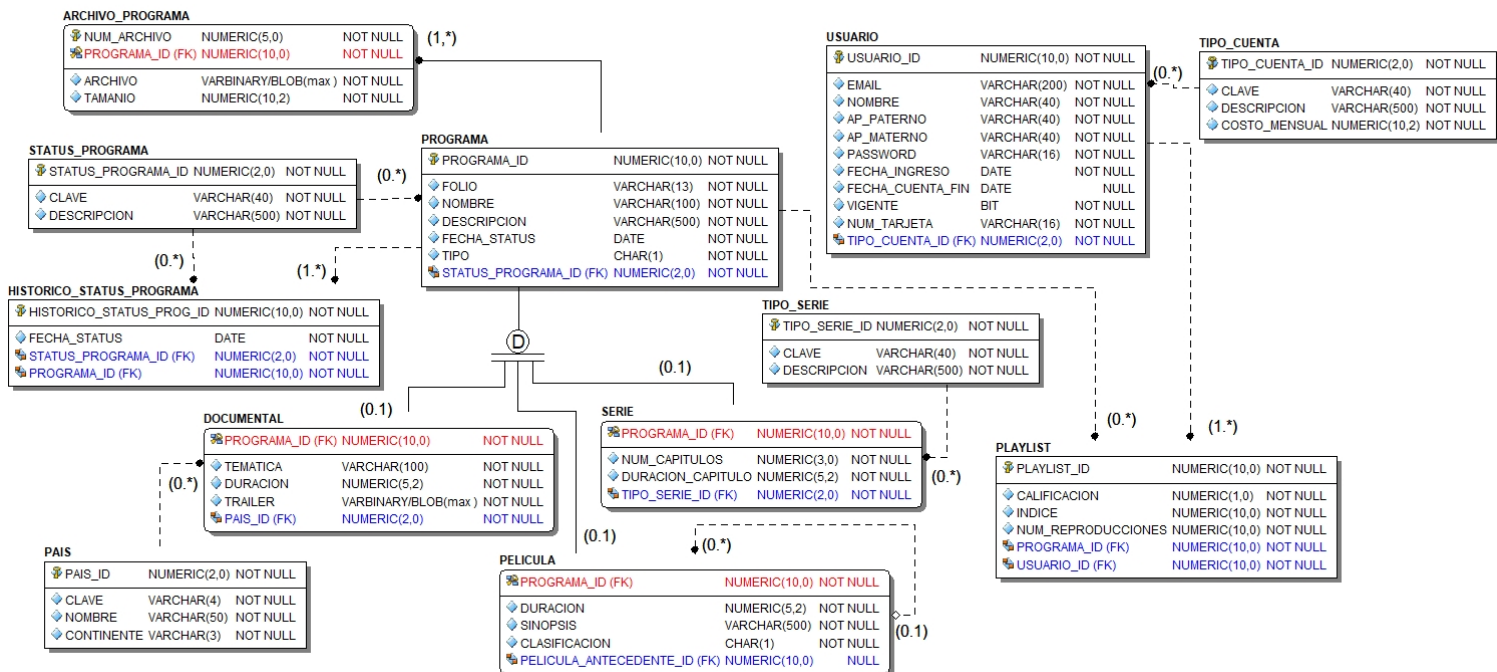
Para las series se almacena número de capítulos, duración del capítulo (todos los capítulos duran lo mismo), y el tipo de serie: A = ACCION, D = Drama, etc., se cuenta con catálogo.

Para poder realizar el streaming de video, cada programa es dividido en una serie de archivos (videos) de hasta 10 MB los cuales son almacenados en la BD. Para cada programa se guarda el número de archivo iniciando en 1, el archivo binario y su tamaño.

Se lleva un registro de todos los clientes con los siguientes datos: nombre, apellidos, email (debe ser único ya que este valor se empleará como username), password, fecha de ingreso, tipo de cuenta: B = BASICA, P = PROFESIONAL, V = VIP. A cada tipo de cuenta se le asigna su costo mensual. Cuando el usuario cancela su cuenta, se registra la fecha fin de su cuenta, y se actualiza una bandera llamada VIGENTE a false. Adicionalmente se almacena el número de tarjeta de crédito para realizar los cargos mensuales.

Finalmente, cada usuario cuenta con un Playlist en el que se guardan todos los "likes" de los programas de un cliente. A cada like se le asigna un puntaje que va de 1 a 5. Para cada entrada del PLAYLIST de cada usuario se le asigna un número consecutivo que inicia en 1.

El diseño global de esta base de datos se muestra en la siguiente figura. Los valores de cardinalidades (1,1) se omiten del modelo para simplificar su representación.



1.1.2. Requerimientos de distribución.

Considere las siguientes reglas que NetMAX ha establecido para realizar el proceso de distribución de su base de datos con el principal objetivo de paralelizar y explotar al máximo las capacidades de sus servidores. La empresa cuenta con 4 nodos disponibles en la que se realizará la distribución.

Para la siguiente tabla, proporcionar los valores para la tercera y cuarta columna. Tomar como ejemplo los valores en azul. En la tercera columna se escribe el nombre global de una PDB elegida al azar que representará a cada uno de los 4 nodos. En la cuarta columna se muestra la cadena S<n> que se emplea para nombrar a cada fragmento: <NOMBRE_GLOBAL>_F<n>_<INICIALES>_S<n>, donde:

- <NOMBRE_GLOBAL> Se refiere al nombre global de la tabla
- F<n> Se refiere al número de fragmento F1, F2, etc.
- <INICIALES> se refieren a las iniciales de los integrantes de equipo, por ejemplo, JRC y ARC. Si se realiza de forma individual, cambiar la primera letra de una de las iniciales por la letra I (Individual), por ejemplo: JRC, IRC.
- S<n> se refiere al número de PDB en un mismo nodo: S1 o S2.

Ejemplos:

Los siguientes nombres corresponden a fragmentos de la tabla USUARIO, que están en cada una de las 4 PDBs

```
USUARIO_F1_JRC_S1
USUARIO_F2_JRC_S2
USUARIO_F3_ARC_S1
USUARIO_F4_ARC_S2
```

Num. nodo	Características	Nombre global del PDB	Sufijo para fragmentos
1	<ul style="list-style-type: none"> • Servidor con capacidad de procesamiento y memoria limitados. 	jrcbd_s1.fi.unam	JRC_S1
2	<ul style="list-style-type: none"> • Servidor con grandes capacidades de procesamiento, memoria y alta disponibilidad. 	jrcbd_s2.fi.unam	JRC_S2
3	<ul style="list-style-type: none"> • Servidor con software especializado para editar videos enfocados al aprendizaje y a la educación. • Su capacidad de procesamiento y memoria se considera como medio • Cuenta con licencias y herramientas de seguridad que permiten cifrar y proteger datos delicados o sensibles. 	arcbd_s1.fi.unam	ARC_S1
4	<ul style="list-style-type: none"> • Servidor con gran capacidad de almacenamiento, en especial para realizar el tratamiento de datos estáticos o históricos que ya no sufren actualizaciones. 	arcbd_s2.fi.unam	ARC_S2

Para el caso de los usuarios, la información se distribuye con base a su tipo de cuenta: VIP (id =3), PROFESIONAL (id = 2) o BÁSICA (id =1). Las cuentas básicas no requieren capacidades de procesamiento y almacenamiento altas. Las cuentas tipo PROFESIONAL se almacenan en el nodo que fue designado para almacenar contenido educativo, y las VIP en el servidor que contiene la mayor capacidad de procesamiento. Por seguridad, los números de tarjeta y el password de todos los usuarios se almacenan en el servidor que cuenta con las herramientas necesarias para realizar su cifrado y correcto manejo. Finalmente, todos los usuarios que tengan su cuenta marcada como no vigente sin importar el tipo de cuenta, se almacenan en el servidor encargado del análisis de datos históricos.

Para el caso de los programas se toma como primer criterio los 2 primeros dígitos del folio. Todos aquellos programas con folio [AA-MZ] se ubicarán en el sitio con capacidades de procesamiento y memoria limitadas. Para los folios restantes: [NA-ZZ] se aplica considerando su estatus. Programas con estatus SATURADO (id=3), programas con estatus DISPONIBLE (id=2) y PROGRAMADO (id=1) en el sitio de capacidades medianas, y programas con estatus NO DISPONIBLE (id = 4) u OBSOLETO (id=5) en el servidor de históricos.

Para realizar el almacenamiento de los datos referentes a documentales, películas y series, se emplean los mismos criterios empleados para almacenar los datos generales del programa.

El histórico de cambios de status de programas se almacenará en el servidor encargado del análisis de datos históricos, no se realiza distribución.

Para el caso de los videos que integran a cada programa, si el archivo tiene un tamaño en MB mayor a 10MB, Los datos y el video se guardarán en el servidor con la mayor capacidad de almacenamiento, de lo contrario se almacenarán en el servidor que contiene software especializado de edición.

Finalmente, para almacenar el playlist se tienen 2 posibilidades: seguir el esquema de fragmentación de PROGRAMA o el de USUARIO (Seleccionar alguno de los 2 criterios).

Con base a las reglas de fragmentación anteriores y las convenciones para nombrar fragmentos, llenar la siguiente tabla que muestra la lista de fragmentos que integrarán a la BDD.

1.1.3. Requerimientos de replicación y copia de datos.

- La tabla `status_programa` conserva su nombre en los 4 nodos ya que se realizará la carga manual en cada uno. No requiere tratamiento alguno.
- Para las tablas `tipo_cuenta`, `pais` y `tipo_serie`, se deberán renombrar por `tipo_cuenta_r_<iniciales>_s<n>`, `pais_r_<iniciales>_s<n>` y `tipo_serie_r_<iniciales>_s<n>` ya que serán replicadas. Más adelante se detallará la estrategia de replicación.
- Agregar estas 3 tablas en la tabla de fragmentación indicando como valor para la expresión algebraica la palabra COPIA MANUAL o TABLA REPLICADA según corresponda.

Nombre del fragmento/Tabla	Expresión algebraica

Continuar con la parte 2 del proyecto.

Nota: Se recomienda ampliamente validar la tabla de fragmentación con el profesor antes de continuar.