

TAREAS. Unidad 1: Almacenamiento de la información

Alumno: Miguel Ángel Brea Rivera

1.Imagina que vas a desarrollar una aplicación para llevar el control de las ventas de ordenadores de una tienda de informática y en principio no vas a utilizar ningún gestor de BD

¿Podrías crear dicha aplicación?

- Si se podría crear dicha aplicación sin usar una base de datos, por ejemplo, si estamos usando lenguaje java en la aplicación podemos usar arrays, listas dinámica o estáticas y ficheros para guardar dicha información y luego poder acceder a dicha información.

¿de qué elemento se dispone en un Sistema Operativo para guardar la información?

- Los Sistemas Operativos disponen de fichero, formularios, carpetas, directorios... para guardar la información necesaria.

¿qué problemas crees que va a plantear este desarrollo?

- A medida que vaya creciendo los datos almacenados en la aplicación, tendrán un grave problema de eficiencia y de mantenimiento en la aplicación para manejar una gran cantidad de datos. Por este motivo, se hace necesario el uso de bases de datos, para definir, manipular o buscar datos específicos dentro de una gran cantidad de información

¿de qué tipos pueden ser los elementos utilizados para guardar la información?

Pon un ejemplo de la información que guardarías en cada uno

- Existen varios tipos de elementos utilizados para guardar información, como por ejemplo varios tipos de ficheros:

Ficheros permanentes que son aquellos que contienen la información relevante de una aplicación. Dentro de estos ficheros se subdivide en: **ficheros maestros**, que es la parte central de la aplicación (ejemplo el ejecutable), es decir su núcleo; en **ficheros constantes** que incluyen datos fijos para la aplicación (por ejemplo, los datos de la app); **ficheros históricos**, que contienen datos que fueron considerados como actuales en un periodo anterior (ejemplo fichero que indican búsquedas de una app).

Otro tipo de fichero son los **temporales**, donde se almacenan información útil para una aplicación. Tienen un periodo corto de existencia. Se subdivide en: **ficheros intermedios**, que almacenan resultados de una aplicación que serán utilizados por otras; **ficheros de maniobras**, que almacenan datos de una aplicación que no puede ser mantenidos en memoria; **ficheros de resultados**, que almacenan datos que van a ser transferidos a un dispositivo de salida.

Todos estos ficheros se encuentran en el almacenamiento secundarios como disco duro, donde el sistema operativo acceden siempre que lo necesiten

2.Crea una tabla con las ventajas e inconvenientes que presentan las distintas formas de organizar los archivos o ficheros para BD.

- Organización Secuencial

Ventajas	Inconvenientes
Aprovechamiento al máximo del soporte de almacenamiento. Al no dejar huecos vacíos. Más compacto.	Para acceder a un dato hay que leer antes todos los datos previos. No accede de forma directa a un dato. Provoca consultas muy lentas.
Se pueden grabar en cualquier tipo de soporte (secuencial o direccionable)	Solo realiza lectura hacia un sentido, hacia delante.
Todos los lenguajes de programación tienen instrucciones para poder trabajar con este tipo de fichero.	Posee una estructura rígida de campos. La posición de campos de cada registro debe ser siempre la misma.
Son muy rápidos para obtener registros contiguos de una base de datos	Es difícil realizar las operaciones de escritura y lectura, necesita de algoritmos mas complejos
	Para actualizarlo, se debe crear un nuevo archivo.
	No se pueden insertar registros entre los que ya están grabados
	No se pueden borrar registros, aunque si se puede marcar para no ser leidos

- Organización Secuencial Encadenada

Ventajas	Inconvenientes
Se mantiene el orden en el que se añadieron los registros y un segundo orden en función de una clave.	No se borran los registros directamente, sino que se marcan como no leído, para ello necesitan una reorganización
Para ordenar no es necesario reorganizar todo el fichero, solo hay que modificar los punteros.	Para añadir registros o cambiar las claves se requiere recalcular los punteros
Las mismas ventajas que poseen la organización secuencial	

Se pueden borrar los registros y después de reorganizar se borran definitivamente.

- Organización Secuencial Indexada

Ventajas	Inconvenientes
El archivo esta siempre ordenado en función a una clave	Para tener un buen uso hay que reorganizar el archivo principal y esta acción es costoso, tener que reescribir de nuevo y de forma ordenada.
La búsqueda de datos es muy rápida	Añadir nuevos registros requiere más tiempo porque se debe de reordenar los índices.
Se puede realizar la búsqueda secuencial en el orden de la clave	
Se puede borrar los registros, aunque es más difícil que otras opciones.	

- Organización directa o aleatoria

Ventajas	Inconvenientes
Acceso muy rápido al no tener que leer datos anteriores.	No se pueden borrar datos (sí marcar para borrado, pero generarán huecos)
La modificación de datos es muy simple.	Las consultas sobre multitud de registros son más lentas
Se permite el acceso secuencial.	
Se permite leer y escribir a la vez.	
Aptos para organizaciones relativas directas, en las que la clave del registro se relaciona con su posición en el archivo.	

3.Cita las ventajas de las BD sobre los sistemas de gestión de datos basados en ficheros.

Ventajas:

- La independencia de los datos, de los programas y de los procesos, permite modificar los datos sin tener que modificar el código de las aplicaciones.

- La integridad de los datos, existe mayor dificultad de perder los datos o de realizar incoherencias con ellos.
- Datos más documentados o detallados, a través de los metadatos que permiten describir la información de la base de datos.
- La necesidad de menor espacio de almacenamiento, debido a una mejor estructuración de los datos.
- Existe una menor redundancia, porque no hace falta tanta repetición de datos. Sólo se indica la forma en la que se relacionan los datos.
- Hay mayor seguridad en los datos. Ya que podemos permitir limitar el acceso a los usuarios. Cada tipo de usuario podrá acceder a un determinado tipo de datos.
- Un acceso a los datos más eficiente. La consecuencia es que la organización de los datos produce un resultado más óptimo en rendimiento.
- Se puede realizar un acceso simultáneo a los datos. Es más fácil controlar el acceso de usuarios.

(...)

4. Teniendo en cuenta los distintos modelos de BD, ¿qué tipo de BD es el más utilizado en la actualidad?

- La base de datos más utilizado en la actualidad son los que utilizan el modelo relacional.

¿qué tipos de BD no se utilizan actualmente?

- Se encuentra en desuso las bases de datos que utilizan el modelo jerárquico o también llamada modelo árbol. La base de datos que utilizan el modelo de red, aunque hay empresa que usan este modelo, tiene una tendencia a ser cada vez menos usada.

¿qué tipo de BD no SQL están apareciendo en la actualidad y en qué tipo de aplicaciones se suelen utilizar?

- Las Bases de Datos no-SQL se diferencia respecto de las Bases de Datos de SQL en los siguientes puntos: No utilizan SQL como lenguaje de consulta; no utilizan estructuras fijas como tablas para el almacenamiento de los datos; no suelen permitir operaciones JOIN; tienen una arquitectura distribuida.

Están apareciendo varios tipos de Bases de Datos no-SQL como son:

- **BD clave-valor:** Es el modelo no-SQL mas popular y más sencillo de usar. Cada elemento está identificado por una llave única, lo que permite la recuperación de la información rápidamente. Algunos ejemplos son Cassandra, BigTable o HBase.

- **BD documentales:** Almacena la información como si fuera un documento, utilizando para ello estructuras simples como son JSON y XML donde se utiliza una clave única para cada registro. Permite realizar búsquedas más avanzadas sobre el contenido del documento. Ejemplo de BD son MongoDB y CouchDB.
- **BD en grafos:** la información se representa como nodos de un grafo y sus relaciones con las aristas del mismo. Su estructura debe estar totalmente normalizada de forma que cada tabla tenga una sola columna y cada relación dos. Este tipo de BD ofrece mayor eficiencia de navegación con respecto a las BD relaciones. Algunos ejemplos son Neo4j, InfoGrid, Virtuoso...
- **BD orientado a objetos:** la información se representa mediante objetos, de la misma forma de como se representa en los lenguajes de programación orientado a objeto POO. Algunos ejemplos de BD son Zope, Gemstone o Db4o.
- Las aplicaciones que usan estos tipos de Bases de Datos no-SQL son aquellas principalmente orientada a las Telecomunicaciones o aquellas que manejan una inmensa cantidad de datos ya sea para investigación o para la actividad privada de cualquier empresa. Por ejemplo, empresas como Netflix usan Cassandra (BD no-SQL del tipo clave-valor) para sus streaming, Facebook usa Cassandra para búsquedas de bandeja, CERN usa un prototipo basado en Cassandra para su proyecto ATLAS, BigTable es usado por Google para sus búsquedas, ...

5. ¿Cuántos modelos conceptuales se pueden tener de una base de datos?

- Tiene un único modelo conceptual, por ejemplo el modelo entidad-relación.

¿y externos?

- En cambio, una base de datos puede tener varios modelos externos, uno para cada grupo de usuario al que son destinado.

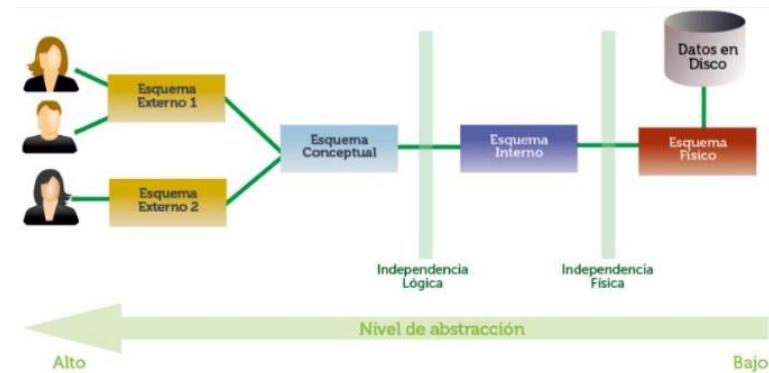
¿qué diferencia hay entre ellos?

- Los modelos conceptuales se centran principalmente en la representación de esquemas sobre las entidades, atributos, relaciones, operaciones de usuarios y restricciones de la información obtenida, ocultando los detalles de las estructuras de almacenamiento. Y los modelos externos se describen varios esquemas externos, donde cada esquema describe la parte de la base de datos que interesa a un grupo de usuario determinados y oculta al resto de grupos el esquema.

Podemos decir que los modelos externos son los mas cercanos al usuario ya que describen los datos o parte de los datos que más le interesan a un grupo de usuario. En cambio, los modelos conceptuales es una

descripción más generalizadas de los datos, ya que van orientados a todos los tipos de usuarios.

6. Dada la imagen siguiente, describe los pasos que sigue el SGBD para interpretar una solicitud de datos de un usuario determinado.



Pasos que sigue el SGBD cuando un usuario solicita datos:

1. El usuario llama al SGBD y hace una instrucción, una llamada a los datos que necesita de la base de datos.
2. El SGBD interpreta esta instrucción o llamada, traduciéndolo a términos del esquema lógico de la base de datos. Accediendo al esquema lógico, comprobando los derechos de acceso y buscando el lugar físico donde se encuentran los datos.
3. El SGBD obtiene el esquema físico.
4. El SGBD traduce la instrucción o llamada a los métodos de acceso del Sistema Operativo que permitirán acceder a los datos requeridos por el usuario.
5. El Sistema Operativo accede a los datos tras traducir las órdenes dadas por el SGBD.
6. Los datos pasan del disco a una memoria intermedia o buffer. En el buffer se va almacenando los datos según se vayan recibiendo.
7. Los datos pasan del buffer al área de trabajo del usuario. Los pasos 6 y 7 se repiten hasta que se envíen toda la información solicitada por el usuario.
8. En caso de haber errores en cualquier momento de este proceso, el SGBD devolverá indicadores en los que manifiesta si ha habido errores o advertencias al área de trabajo del usuario.

¿Qué significa el nivel de abstracción?

- El nivel de abstracción hace referencia a la visión de los datos desde diferentes puntos de vistas, ya sean externa o física. La visión externa sería los datos que obtienen los usuarios de una aplicación y la vista física

es como realmente es la información de cómo están almacenados los datos. Es importante la abstracción de los datos, a mayor abstracción mayor independencia física de los datos y más protegidos están de ser manipulados incorrectamente. En los niveles de abstracción podemos diferenciar:

Nivel físico, donde se representa la forma en que están almacenados los datos. La visión de este nivel solo lo requieren los administradores para poder gestionar más eficientemente los datos. En este esquema encontramos archivos, directorios, unidad de discos, servidores, ...

Nivel conceptual, es un esquema teórico de los datos en el que los datos se organizan de forma que sean reconocible al mundo real y también se representa sus relaciones. Este nivel lo que hace es modelar un problema real a su forma correspondiente en el ordenador. Es el primer paso, para crear una base datos. Lo realizan diseñadores y analistas.

Nivel externo, se trata de una visión de los datos que poseen los usuarios y usuarios finales. Esta visión es la que se obtiene a través de las aplicaciones. Cada aplicación produce un esquema externo diferente según sus intereses.

7. ¿Qué funciones realiza un SGBD? Describe cada una de ellas.

- **Función de definición o descripción:** Permite al diseñador de la base de datos crear la estructura apropiada para integrar adecuadamente los datos. Permite definir las tres estructuras de la base de datos (interna, conceptual, externa). Se realiza mediante el lenguaje de descripción de datos DDL donde se definen la estructura de datos, las relaciones entre los datos y las reglas que han de cumplir los datos.
- **Función de manipulación:** Permite modificar y utilizar los datos de la base de datos. Mediante el lenguaje de modificación de datos DML se puede añadir, eliminar, modificar o buscar datos. Aunque se suele distinguir en la parte de búsqueda de datos con el uso del lenguaje de consulta de datos DQL.
- **Función de control:** con esta función los administradores tienen mecanismos para controlar las visiones de los datos permitidos a cada usuario, además de proporcionar elementos de creación y modificación

de esos usuarios. Incluimos en esta función las tareas de copias de seguridad, carga de ficheros, auditorías, protección ante ataques externos, configuración del sistema, ... El lenguaje que implementa todas estas funciones es el lenguaje de control de datos DCL.

8. En función de la ubicación de la información

¿qué tipos de almacenamiento podemos tener?

- Los almacenamientos más utilizados para las bases de datos son: Disco SATA, discos SCSI, disco RAID, los sistemas NAS y los sistemas SAN.

¿qué ventajas e inconvenientes presenta cada uno de ellos?

Almacenamiento	Ventajas	Inconvenientes
Disco SATA	Proporcionan mayores velocidades, mejor aprovechamiento cuando hay varias unidades, mayor longitud del cable de transmisión de datos y capacidad para conectar unidades al instante. Un precio bastante asequible.	Su velocidad es mucho menor a los demás dispositivos. En caso de usar mucha información con regularidad, con el paso del tiempo se va a notar errores. Es clave la instalación de estos disco en un sistema RAID para evitar inconvenientes.
Disco SCSI	Son interfaces preparadas para discos duros de gran capacidad de almacenamiento y velocidad de rotación. Cuentan con mayor duración en su ciclo de vida. Garantiza una mayor confiabilidad.	Para el uso de grandes cantidades de datos o información no es eficiente el uso de un solo disco SCSI.
Disco RAID	Para obtener una ampliación de capacidad, mejoras en velocidad y seguridad de almacenamiento	No se facilita la recuperación de datos y tampoco se asegura la protección de los datos.
Sistemas NAS	Permiten compartir la capacidad de almacenamiento de un ordenador (Servidor) con ordenadores personales o servidores clientes a través de una red.	En caso de usar una gran cantidad de datos, es necesario para ser eficiente un gran desembolso económico en hardware. La conexión a internet es otro de los problemas que pueden presentar, la velocidad contratada juega un papel importante. La seguridad del sistema dependerá de nuestras manos. Es decir, que el sistema operativo este actualizado y el software en la última versión.
Sistemas SAN	Permite que los recursos de almacenamiento estén disponibles para varios servidores en una red de área local o amplia	El coste de adquisición de este sistema es mucho más caro, porque se trata de una arquitectura compleja.

9. ¿Cuáles son los componentes de un SGBD?

- Los lenguajes de BD, el diccionario de datos, el gestor de la BD, el usuario de la BD y las herramientas de la BD.

¿qué función realiza cada uno de ellos?

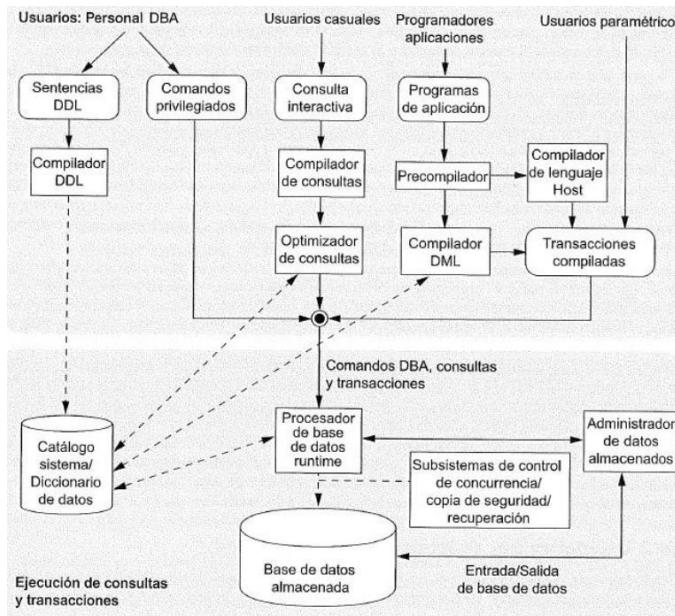
- Lenguaje de BD: a través de los lenguajes se pueden especificar los datos que componen la BD, su estructura, relaciones, reglas de integridad, control de acceso, características físicas y vistas externas de los usuarios.
- Diccionario de datos: Es el lugar donde se deposita la información sobre la totalidad de los datos que forman la BD. Contiene las características lógicas de las estructuras que almacenan los datos, su nombre, descripción, contenido y organización.
- Gestor de Base de Datos: software encargado de garantizar el correcto, seguro, íntegro y eficiente acceso y almacenamiento de los datos.
- Usuario de la Base de Datos: existen diferentes perfiles de usuarios, uno de ello es el administrador encargado de la gestión o administración de los datos, después encontramos otros tipos de usuarios donde tienen diferentes accesos a los datos y diferentes necesidades de usar un dato u otro.
- Herramientas de Base de Datos: conjunto de aplicaciones que permiten a los administradores la gestión de la BD, de los usuarios y permisos, generadores de formularios, informes, interfaces gráficas, generadores de aplicaciones, etc.

Describe y comenta la imagen

- En la imagen se representa los distintos tipos de usuarios que utilizan un sistema gestor de base de datos, por ejemplo:
- El personal DBA son los administradores de la base datos, entonces a través del SGBD son capaces de realizar sentencias DDL para definir catálogo y diccionario de datos que se almacenarán en la base de datos. También realizarán comandos privilegiados para poder realizar consultas y transacciones de datos, usando el procesador de base de datos runtime, teniendo en cuenta que además de acceder a la base de datos se generará paralelamente copias de seguridad.
- Podemos observar otros tipos de usuarios como los casuales que realizarán consultas interactivas mediante la compilación de la consulta y optimizando dichas consultas el SGBD ofrecerá los datos solicitados por el usuario.
- Por otro lado, hace mención a los programadores de aplicaciones que mediante la precompilación de lenguajes de programación obtienen las transacciones compiladas que necesite o directamente usando solo el

lenguaje de Manipulación de datos son capaces de modificar la base de datos.

- Los usuarios paramétricos solo acceden a las transacciones compiladas sin necesidad de programar aplicaciones ni nada por el estilo, solo usan una aplicación, piden los datos que necesitan a través de esa aplicación y a través del SGBD la aplicación le ofrece los resultados de dicha consulta.



¿Qué permiten las BD distribuidas?

- La Base de Datos distribuidas tiene una estructura operacional que se denomina también multi-servidor que ocurre cuando los clientes acceden a datos situados en más de un servidor. Por tanto, el cliente no sabe si los datos están en uno o más servidores, ya que el resultado es el mismo independientemente de dónde se almacenen los datos. Podemos decir que en este tipo de estructura de BD hay un servidor de aplicaciones que es el que recibe las peticiones y el encargado de traducirlas a los distintos servidores de datos para obtener los resultados.

Enumera sus ventajas e inconvenientes.

Ventajas	Inconvenientes
El acceso y procesamiento de los datos es más rápido ya que existen más nodos que comparten cargas de trabajo.	El mantenimiento y el control de la red puede resultar costoso.
Si un nodo deja de funcionar, la estructura podría seguir funcionando, es decir, existe una cierta tolerancia a fallos.	La probabilidad de fallos de seguridad es creciente si no se toma precauciones.
Costes inferiores si lo comparamos con las bases de datos centralizadas	Se debe crear un sistema de coordinación de los nodos para que sean eficientes

Se adapta más a la estructura de las organizaciones,	El control de concurrencia y los mecanismos de recuperación son muchos más complejos que un sistema centralizado.
Aunque los nodos están interconectados, existen una independencia local.	Constituye una forma de sobrecarga que no surge en el sistema centralizado
Podemos acceder a la información almacenada en distintos lugares.	El intercambio de mensajes y el cómputo adicional necesario para conseguir la coordinación entre los distintos nodos.
	Dada la complejidad del procesamiento entre nodos es difícil asegurar la corrección de los algoritmos, el funcionamiento correcto durante un fallo o la recuperación.

11. Considera la siguiente relación, ¿qué fragmentos se obtendrían al realizar una fragmentación horizontal por ESCUELA?

Jno	NOMBRE	NOTA	ESCUELA
J1	LUIS YANEZ	8	CIME
J2	ERIKA QUIROZ	8	CIME
J3	DANIEL MURILLO	9	EISIC
J4	MARIA JOSE MENDEZ	10	EISIC
J5	ADONIS PABON	9	EISIC

- Las tablas obtenidas tras la fragmentación horizontal por escuela:

Jno	NOMBRE	NOTA	ESCUELA
J1	LUIS YANEZ	8	CIME
J2	ERIKA QUIROZ	8	CIME

Jno	NOMBRE	NOTA	ESCUELA
J3	DANIEL MURILLO	9	ESIC
J4	MARIA JOSE MENDEZ	10	ESIC
J5	ADONIS PABON	9	ESIC

12. Considera las siguientes relaciones, ¿qué fragmentos principales y derivados se obtendrían al realizar una fragmentación horizontal por ESCUELA?

ID	MATERIA	ESCUELA	CRÉDITOS	NIVEL
1	Análisis Matemático	EISIC	4	1
2	Sistemas Operativos	EISIC	6	3
3	Programación II	CIME	6	2
4	Tecnología Eléctrica	EISIC	4	2
5	Técnicas de Aprendizaje	CIME	4	1
6	Dibujo Mecánico	CIME	6	3

Jno	NOMBRE	MATERIA	NOTA
J1	LUIS YANEZ	Sistemas Operativos	8
J2	ERIKA QUIROZ	Dibujo Mecánico	8
J3	DANIEL MURILLO	Técnicas de Aprendizaje	9
J4	MARIA JOSE MENDEZ	Ánalisis Matemático	10
J5	ADONIS PABÓN	Programación II	9

- Fragmentos principales tras la fragmentación por escuela serían lo siguientes:

ID	MATERIA	ESCUELA	CREDITOS	NIVEL
1	Análisis Matemático	EISIC	4	1
2	Sistema Operativos	EISIC	6	3
4	Tecnología Eléctrica	EISIC	4	2

ID	MATERIA	ESCUELA	CREDITOS	NIVEL
3	Programación II	CIME	6	2
5	Técnicas de Aprendizajes	CIME	4	1
6	Dibujo Mecánico	CIME	6	3

- Los fragmentos derivados se relacionan con la tabla principal a través de la materia, por tanto obtendremos las siguientes tablas:

ID	Jno	NOMBRE	MATERIA	NOTA	ESCUELA	CRED.	NIVEL
1	J4	MARIA JOSE MENDEZ	Análisis Matemático	10	EISIC	4	1
2	J1	LUIS YANEZ	Sistema Operativos	8	EISIC	6	3
4			Tecnología Eléctrica		EISIC	4	2

ID	Jno	NOMBRE	MATERIA	NOTA	ESCUELA	CRED.	NIVEL
3	J5	ADONIS PABON	Programación II	9	CIME	6	2
5	J3	DANIEL MURILLO	Técnicas de Aprendizajes	9	CIME	4	1
6	J2	ERIKA QUIROZ	Dibujo Mecánico	8	CIME	6	3

13. Considera la relación del ejercicio 11. Realiza una fragmentación vertical por nota y por nombre/escuela.

Jno	NOMBRE	NOTA	ESCUELA
J1	LUIS YANEZ	8	CIME
J2	ERIKA QUIROZ	8	CIME
J3	DANIEL MURILLO	9	EISIC
J4	MARIA JOSE MENDEZ	10	EISIC
J5	ADONIS PABON	9	EISIC

- Las tablas obtenidas después de las fragmentaciones vertical por notas y por nombre de escuela son:

La columna Jno se mantiene porque se trata de la clave primaria.

Jno	NOTA	ESCUELA
J1	8	CIME
J2	8	CIME
J3	9	EISIC
J4	10	EISIC
J5	9	EISIC

