

UD1-Sistemas de almacenamiento

2.- Ficheiros

DAM1 – BASES DE DATOS

ASIR1 – XESTIÓN DE BASES DE DATOS

ORIENTACIÓN AOS PROCESOS

- Xurde da automatización dos procesos manuais que se realizaban en cada departamento das organizacións.
- Os datos de cada ficheiro están fortemente ligados ao programa que os procesa.
- Os procesos e os datos almacénanse en citas magnéticas. Para o procesamento cárganse os procesos e os datos no equipo. Despois do procesamento os datos resultantes vólcanse de novo en cinta.
- Cada proceso manexa os seus propios ficheiros:
 - Ficheiros con datos de entrada, cos datos a procesar.
 - Ficheiros con datos de saída, cos datos resultantes de realizar o procesamento.
- Cada ficheiro contén datos relacionados cun determinado proceso.
- Os datos de cada ficheiro son independentes doutros ficheiros e doutros procesos.

SISTEMAS DE FICHEIROS

REXISTROS

- Nos ficheiros os datos organízanse en forma de rexistros.
- Cada rexistro agrupa un conxunto de datos relacionados relativos ao mesmo obxecto ou elemento.
- Cada dato correspóndese cunha categoría que lle da significado.
- Exemplo:

73564765M	Javier	Barquín Arce	C/ Alta, 234	918342156
-----------	--------	--------------	--------------	-----------

- Un rexistro contén os datos dun cliente: os datos de Javier Barquín.
- Cada dato corresponde a un campo ou atributo: o dni, o nome, o teléfono, ...

SISTEMAS DE FICHEIROS

REXISTROS

Rexistros de tamaño fixo:

- Todos os rexistros teñen o mesmo tamaño.
 - Todos os rexistros teñen os mesmo campos.
 - Cada campo ten un espazo reservado en memoria.
- O valor dun campo non pode exceder o tamaño reservado.
- Se un campo non ten valor ou ten un valor pequeno ocupará igualmente o tamaño máximo: desperdicio de espazo de almacenamento.
- É fácil calcular a posición dun rexistro n :
bastará sumar $n * \text{tamaño}$ á posición inicial.

SISTEMAS DE FICHEIROS

REXISTROS

Rexistros de tamaño variable:

- Os rexistros poden ter tamaños distintos.
- O tamaño dun campo será o que precise o seu valor.
- Afórrase espazo de almacenamento.
- Non podemos determinar onde empeza e remata cada rexistro: é necesario usar algún sistema para identificalo (por exemplo marcas de fin de rexistro).
- Coñecer a posición en memoria de cada rexistro é complexo.
- Cambios nun rexistro esixen reasignar memoria.

ESTRUTURA DOS FICHEIROS

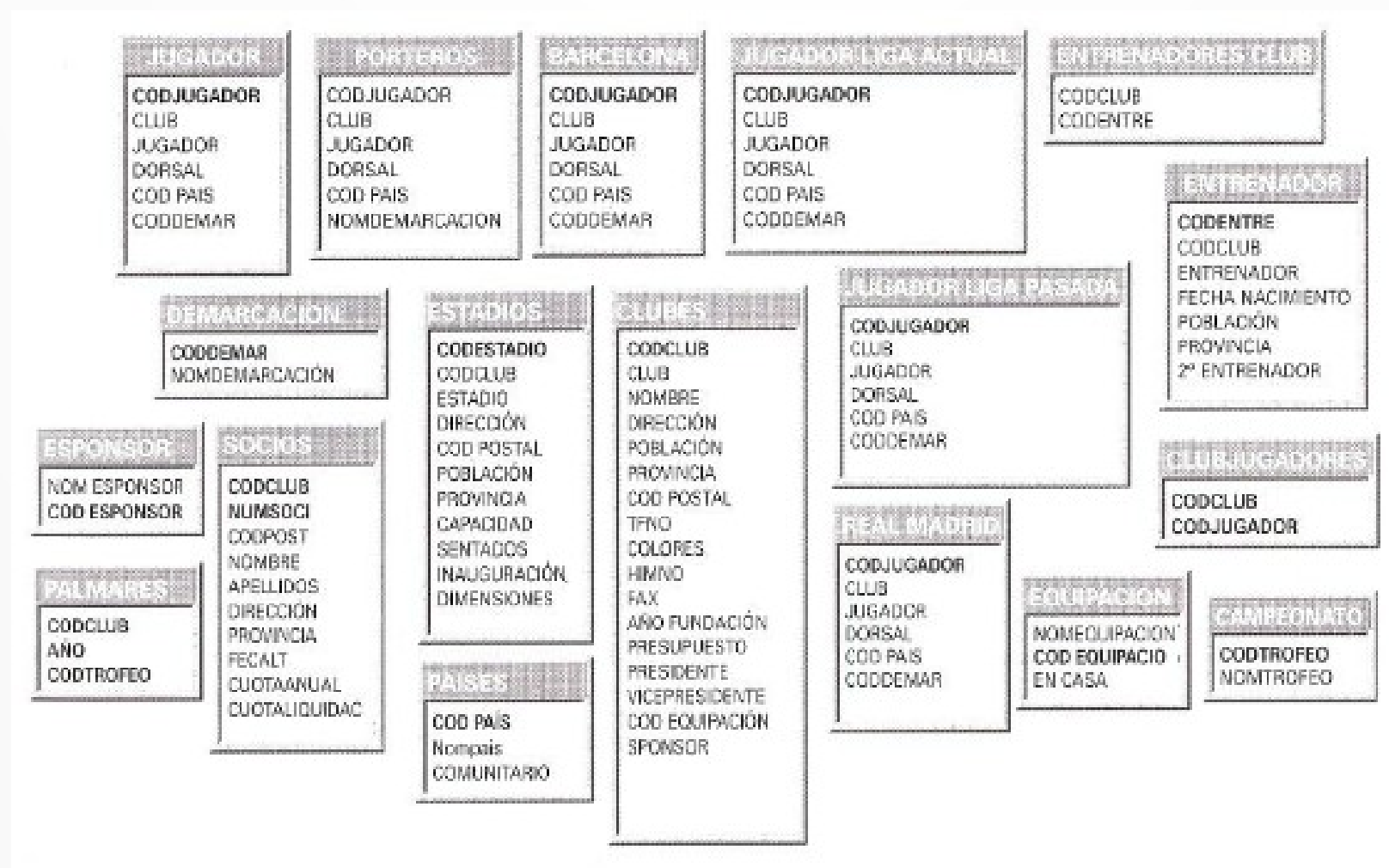
- Existen distintas formas de representar os rexistros en ficheiros, a máis habitual é en forma de táboa:
 - Cada fila representa un rexistro.
 - Cada columna representa un campo.

FICHERO DE CLIENTES					
Campos:	DNI	NOMBRE	APELLIDOS	DIRECCION	TELEFONO
Registro 1:	73564765M	Javier	Barquín Arce	C/ Alta, 234	918342156
Registro 2:	56558765W	Luis	Gómez de Miguel	Avda. de Castilla, 2A	956235567
Registro 3:	13874521M	María Belén	Márquez Ruiz	C/ <u>Floranes</u> , 2	568732212
Registro 4:	75675317R	Carmen	Rodríguez Mata	Paseo Pereda, 123	942665544

UD1 Sistemas de almacenamiento

SISTEMAS DE FICHEIROS

EXEMPLO



INCONVENIENTES DOS SISTEMAS DE FICHEIROS

- **Redundancia:** Os datos que se utilicen en varios procesos van estar repetidos en varios ficheiros.
- **Custo de almacenamento elevado:** Cada dato é almacenado varias veces, cada copia ocupa espazo no disco.
- **Tempos de procesamento elevados:** O tratamento dos datos é secuencial. O procesamento dun só dato implica percorrer ficheiros completos.
- **Modificación custosa:** Se un dato cambia debe modificarse en todos os ficheiros nos que estea.
- **Inconsistencias:** Prodúcese unha inconsistencia se un dato se repite en varios ficheiros con distinto valor. Xeralmente é debido a que un cambio non se reproduciu en todas as copias do dato.

SISTEMAS DE FICHEIROS

TIPOS SEGUNDO O SEU USO

- **Mestres:** Son os principais sistemas de información e conteñen os datos fundamentais. Raramente cambian.
- **Constantes:** Conteñen información que apenas varía ao longo do tempo.
- **Históricos:** Almacenas copias dos ficheiros mestres que se realizan periodicamente
- **De movemento:** Almacenan os cambios que se farán nos ficheiros mestres. Elimínanse despois de facer eses cambios.
- **De manobra:** Archivos auxiliares empregados polo software que xestiona os datos. Destrúense en canto rematan as aplicacións.

SISTEMAS DE FICHEIROS

TIPOS DE FICHEIRO SEGUNDO A FORMA DE ACCESO AOS REXISTROS

- **Secuencial:** Os rexistros están ordenados secuencialmente na orde na que se rexistraron. Para acceder a un rexistro é necesario percorrer todos os rexistros anteriores.
- **Secuencial encadeado:** Son ficheiros secuenciais xestionados usando punteiros. Un punteiro é un datos especial que contén o enderezo doutro rexistro no ficheiro.
- **Directo ou aleatorio:** Permiten acceder directamente a calquera rexistro do ficheiro, para o que é preciso coñecer a posición na que se atopa. A posición pode indicarse en bytes ou mediante unha clave.
- **Indexado:** Os rexistros do ficheiro están agrupados en bloques de rexistros ordenados. Para acceder aos rexistros utilízase un índice en forma de táboa, que relaciona a primeira clave de cada bloque coa posición na que se atopa. Unha vez localizado o bloque, percórrense secuencialmente os seus rexistros ata localizar o buscado.
- **Dinámico:** Permite o acceso directo ou por índice a algún rexistro, e a partir dese accédese aos demais de forma secuencial.

FICHEIRO SECUENCIAL

- Os rexistros escríbense no dispositivo de almacenamento en posicións fisicamente contiguas, sen deixar ocos entre eles, na mesma orde na que se introduciron.
- Este tipo de ficheiros poden empregar dispositivos ou soportes de acceso secuencial ou non direccionables, como cintas de almacenamento magnético de datos.

Dirección de memoria	Marca de borrado	Nombre	Apellidos	Teléfono
1200		Alfredo	Bárcena	768334472
1300	*	Isabel	De los Ríos	987335612
1400		Carmen	Sierra	955347612
1500		Fernando	Ruiz	674992455
1600		Juan Carlos	Abad	573982277

FICHEIRO SECUENCIAL

- **Operacións:**

- **Percorrido do ficheiro:** Os rexistros percórrense secuencialmente empezando polo principio do ficheiro e ata chegar ao final.
- **Acceso a un rexistro:** Require crear un proceso no que se percorra o ficheiro e se compare o valor do campo a localizar co valor do mesmo campo dos rexistros percorridos.
- **Consulta dun rexistro:** Require primeiro acceder ao rexistro. Despois a lectura é directa.
- **Inserción dun rexistro:** Só pode facerse ao final do ficheiro. Non se poden inserir novos rexistro entre os xa existentes.
- **Borrado dun rexistro:** Require primeiro acceder ao rexistro. Despois o rexistro non se pode eliminar fisicamente, se non que se marcará como borrado (borrado lóxico). Os rexistros coa marca de borrados non se teñen en conta nas outras operacións.
- **Modificación dun rexistro:** Non é posible a modificación. No seu lugar procederase a borrar o rexistro e inserilo de novo, co campo modificado.
- Para **manter ordenado e compacto** o ficheiro, é necesario facer periodicamente operacións de mantemento: copiar os rexistros non borrados noutra posición de memoria.

FICHEIRO SECUENCIAL

- Vantaxes:
 - Acceso rápido a un bloque de rexistros contiguos
 - Non se desperdicia espazo de almacenamento no dispositivo, porque non hai ocos
 - Pódese usar calquera tipo de rexistro: lonxitude fixa, variable ou indefinida.

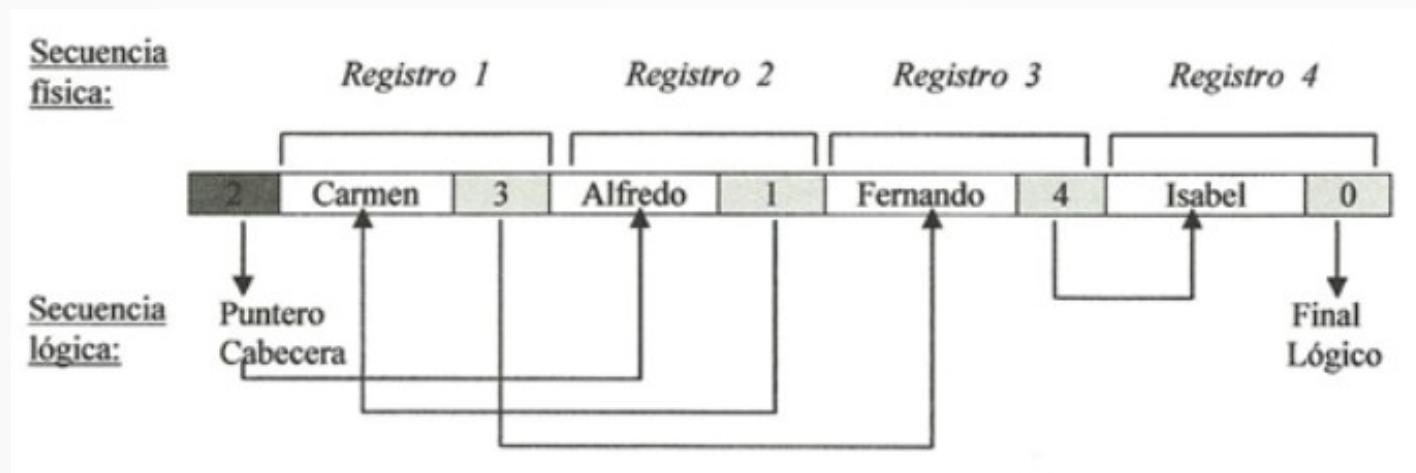
SISTEMAS DE FICHEIROS

FICHEIRO SECUENCIAL ENCADEADO

- Son ficheiros de organización secuencial xestionados mediante punteiros.
- Un punteiro é un campo adicional que almacena a dirección doutro rexistro do ficheiro. Usando o punteiro podemos acceder a ese rexistro.
- O uso de punteiros permite percorrer os rexistros nunha orde distinta da orde na que foron rexistrados, establecendo unha secuencia lóxica.
- Dado un rexistro, chamaremos **rexistro seguinte** a aquel ao que se accede usando o punteiro almacenado no primeiro.
- Os rexistros ficarán ordenados mediante a orde que establece a secuencia de punteiros.
- A orde xeralmente está relacionada co valor dalgún outro campo.

FICHEIRO SECUENCIAL ENCADEADO

- Utiliza dúas marcas especiais:
 - **Punteiro cabeceira**: punteiro adicional que indica onde empeza a secuencia.
 - **Marca de final de ficheiro**: indica onde remata a secuencia.



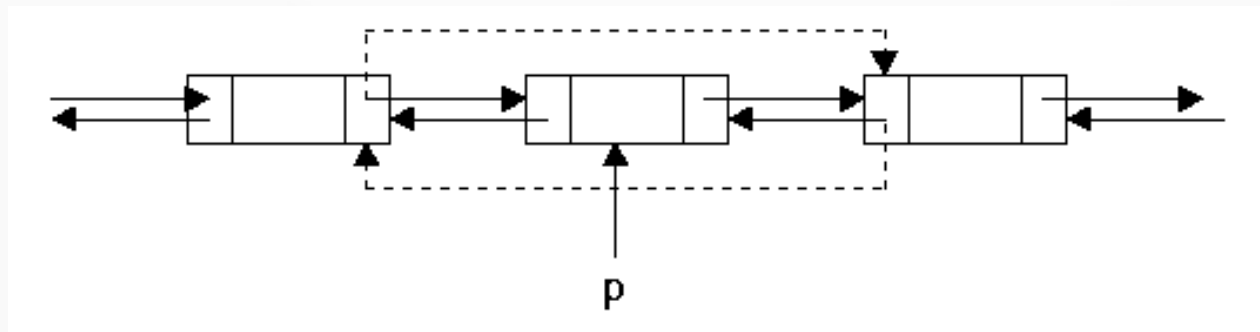
FICHEIRO SECUENCIAL ENCADEADO

- **Operacións:**

- **Percorrido do ficheiro:** Os rexistros percórrense secuencialmente utilizando os punteiros, empezando polo punteiro cabeceira e rematando na marca de final de ficheiro.
- **Acceso a un rexistro:** Require crear un proceso no que se percorra o ficheiro e se compare o valor do campo a localizar co valor do mesmo campo dos rexistros percorridos.
- **Consulta dun rexistro:** Require primeiro acceder ao rexistro. Despois a lectura é directa.
Tipicamente faranse consultas polo criterio de ordenación, polo que bastará con seguir o punteiro ao rexistro seguinte.
- **Inserción dun rexistro:** A inserción física só pode facerse ao final do ficheiro. Para inserir un rexistro entre os xa existentes, inserirase ao final e modificaranse os punteiros para manter a orde.
- **Borrado dun rexistro:** Require primeiro acceder ao rexistro. Despois modificarase o valor do punteiro do rexistro anterior para que apunte ao rexistro seguinte ao que se quere eliminar. O borrado, por tanto, é lóxico.
- **Modificación dun rexistro:** Non é posible a modificación. No seu lugar procederase a borrar o rexistro e inserilo de novo, co campo modificado.
- Para **manter ordenado e compacto** o ficheiro, é necesario facer periodicamente operacións de mantemento: copiar os rexistros non borrados noutra posición de memoria e actualizar os punteiros.

FICHEIRO SECUENCIAL DOBREMENTE ENCADEADO

- Variación dos ficheiros encadeados no que cada rexistro mantén dous punteiros:
 - Ao rexistro seguinte.
 - Ao rexistro anterior.
- Permite percorridos nos dous sentidos da ordenación.
- Permite accesos máis rápidos: usaranse os punteiros no sentido máis conveniente.
- Facilita algunhas operacións, por exemplo o borrado.
- As actualizacións requiren actualizar ambos punteiros.

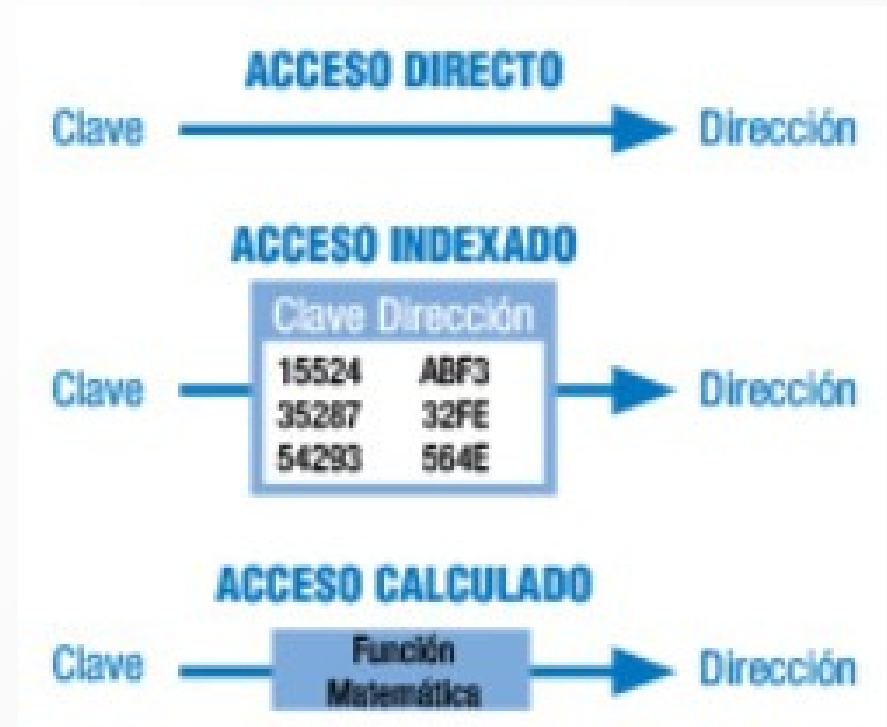


FICHEIRO DE ACCESO DIRECTO OU ALEATORIO

- Cada rexistro ten un campo adicional **chave**, que permite localizalo fisicamente no almacenamento.
- Usando a chave pódese recuperar a posición física do rexistro no almacenamento, e acceder a el directamente.
- Os rexistros poden almacenarse en calquera posición física do disco: acceso aleatorio.
- Adoita usarse con sistemas de almacenamento de memoria masiva de acceso directo, como discos magnéticos.

FICHEIRO DE ACCESO DIRECTO OU ALEATORIO

- A posición física do rexistro obtense aplicando un transformación específica á chave:
 - **Acceso directo:** A chave contén a dirección de memoria onde está o rexistro.
 - **Acceso indexado:** Utiliza unha táboa que relaciona índices con direccións de memoria. A chave contén un índice que se usa para recuperar na táboa a dirección de memoria onde está o rexistro.
 - **Acceso calculado:** Utiliza unha función matemática. A chave contén un valor que ao serlle aplicada dita función permite obter a dirección de memoria onde está o rexistro.



FICHEIRO DE ACCESO DIRECTO OU ALEATORIO

- Outras características:
 - Posicionamento inmediato.
 - Rexistros de lonxitude fixa.
 - Os rexistros bórranse colocando un cero na posición que ocupan.
 - Permiten o uso de algoritmos de compactación de ocos.
 - Na creación do ficheiro establécese o seu tamaño máximo, é dicir, o número máximo de rexistros que pode conter.
 - Úsanse cando o acceso aos datos dun rexistro sempre se fai coa mesma chave e a velocidade de acceso a un rexistro é o que máis nos importa.
 - Permiten a actualización dos rexistros no mesmo ficheiro, sen necesidade de copialo.
 - Permiten actualizar procesos en tempo real.

FICHEIRO DE ACCESO DIRECTO OU ALEATORIO

- **Operacións:**
 - **Percorrido do ficheiro:** Percorrerase o índice secuencialmente e accederase a cada rexistro directamente.
 - **Acceso a un rexistro:** Acceso directo a partir da súa chave usando o índice.
 - **Consulta dun rexistro:** Require primeiro acceder ao rexistro. Despois a lectura é directa.
 - **Inserción dun rexistro:** Se o índice está cheo non se permite inserir máis elementos. Se hai sitio, insírase o elemento en memoria e almacénase no índice, na posición correspondente, a súa chave e a súa posición en memoria.
 - **Borrado dun rexistro:** Modifícase o índice para que a posición do elemento con esa chave tome valor cero.
 - **Modificación dun rexistro:** Require primeiro acceder ao rexistro. Despois modifícanse os campos que corresponda.

FICHEIRO INDEXADOS

- Os rexistros do ficheiro gárdanse secuencialmente en memoria ordenados por un dos seus campos (**chave**).
- O ficheiro divídese lóxicamente en **bloques** de rexistros. Todos os bloques teñen o mesmo tamaño.
- Úsase un **índice** en forma de táboa, que relaciona a primeira chave de cada bloque coa posición en memoria na que se atopa.
- Usando o índice pódese acceder de forma directa ao primeiro rexistro de cada bloque.
- Unha vez localizado o bloque, o acceso aos rexistros será secuencial.



SISTEMAS DE FICHEIROS

FICHEIRO INDEXADOS

- Cada ficheiro consta de 3 áreas:
 - **Área primaria:** Contén os rexistros ordenados secuencialmente pola chave e agrupados en bloques.
 - **Área de índices:** Créaa o sistema no momento de almacenar os datos.
Cada rexistro contén dous campos:
 - ♦ O valor da chave do primeiro rexistro dun bloque.
 - ♦ A posición de memoria do primeiro rexistro do bloque.
 - **Área de overflow ou de excedentes:** Contén os rexistros que se engadiron ao ficheiro despois da súa creación. Os rexistros non están ordenados pola chave e a súa chave pode intercalarse coas da área primaria.

FICHEIRO INDEXADOS

- **Operacións:**
 - **Percorrido do ficheiro:** Non é fácil facer un percorrido ordenado dos rexistros, debido á área de *overflow*.
 - **Acceso a un rexistro:** Pode requirir 2 ou 3 pasos:
 - Utilizárase o índice para acceder directamente ao primeiro rexistro do seu bloque.
 - Realízase un percorrido secuencial dos rexistros do bloque ata atopar o rexistro.
 - Se non se atopa, realízase un percorrido secuencial dos rexistros da área de *overflow* ata atopalo.
 - **Consulta dun rexistro:** Require primeiro acceder ao rexistro. Despois a lectura é directa.
 - **Inserción dun rexistro:** Insérase ao final da área de *overflow*. Non se permiten insercións na área primaria despois da creación.
 - **Borrado dun rexistro:** Require primeiro acceder ao rexistro. Despois o rexistro non se pode eliminar fisicamente, se non que se marcará como borrado (borrado lóxico).
 - **Modificación dun rexistro:** Non é posible a modificación. No seu lugar procederase a borrar o rexistro e inserilo de novo, co campo modificado.
 - Para **manter ordenado e compacto** o ficheiro, é necesario facer reorganizar o arquivo cando hai moitos borrados ou se a área de *overflow* é grande.