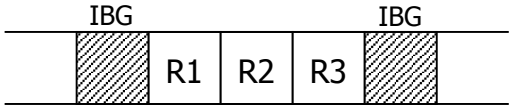
جلسه دوم ذخیره و بازیابی اطلاعات :

بلاک بندی (Blocking):

بلاک واحد رد و بدل کردن اطلاعات بین حافظه جانبی و حافظه اصلی توسط سیستم فایل است. البته در یک عمل I/O ممکن است چندین بلوک یکباره خوانده یا نوشته شوند. از نظر برنامه پردازشگر، فایل مجموعه‌ای از رکوردها با ساختار مشخص است ولی از نظر سیستم فایل، یک فایل از تعدادی بلاک تشکیل یافته است. نمایش ساده یک بلوک:



ضریب بلاک بندی:

به تعداد رکورد های موجود در هر بلاک ضریب بلاک بندی گفته می شود و آن را با BF (مخفف Blocking Factor) نمایش می دهیم. مابین بلاک ها یک فضای بلا استفاده (GAP) وجود دارد که باعث هدر رفتن فضای ذخیره سازی می گردد.

تعداد رکوردها را با n، تعداد بلوک ها را با b، اندازه هر بلوک را با B و اندازه ی هر رکورد را با R نمایش می دهیم. بنابراین داریم:

$$BF = \frac{B}{R}$$

بلاک بندی در نوار توسط کاربر انجام گرفته و اندازه آن می تواند تغییر کند.

«GAP» بین بلوک ها در نوار جهت رسیدن سرعت هد به سرعت حس و یا توقف هد مورد نیاز است.

بلاک بندی در دیسک:

بلاک در دیسک می تواند یک سکتور یا ترکیبی از چند سکتور سخت افزاری، یک شیار یا بخشی از یک شیار باشد. یک بلوک را نمی توان بین دو یا چند شیار تقسیم کرد.

شیارهای دیسک را می توان بر حسب سکتورها یا بر حسب بلوک ها تقسیم بندی کرد، تقسیم بندی بلوکی توسط کاربر و یا سیستم عامل انجام می پذیرد. بلوک واحد رد و بدل اطلاعات بین دیسک و حافظه است و بلوک ها می توانند طول ثابت یا متغیری داشته باشند که بستگی به نیاز طراح فایل و قابلیت های سیستم عامل دارد.

بلوک ها را مشابه سکتورها می توان رکوردهای فیزیکی در نظر گرفت. بلوک ها طوری سازماندهی می شوند که تعداد ثابتی از رکوردهای منطقی را نگهداری می کنند.

تکنیک های بلاک بندی:

۱ـ بلاک بندی رکوردها با طول ثابت و یکپاره

۲ـ بلاک بندی رکوردها با طول متغیر و یکپاره

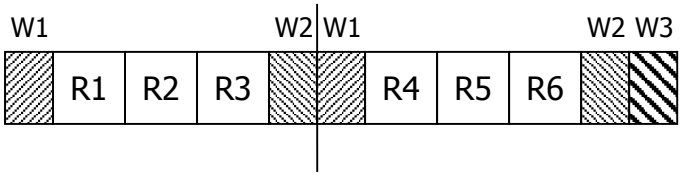
۳ـ بلاک بندی رکوردها با طول متغیر و دوپاره

در هر کدام از این سه مورد که بررسی کنیم باید دو مسئله را مد نظر داشته باشیم:

۱ـ فاکتور بلاک بندی؛ یعنی تعداد رکورد موجود در هر بلاک.

۲ـ حافظه هرز

روش اول:



$W1 = G$ = گپ بین بلاک ها

$W2$ = حافظه هرز ناشی از جا نگرفتن رکورد اضافی در بلاک

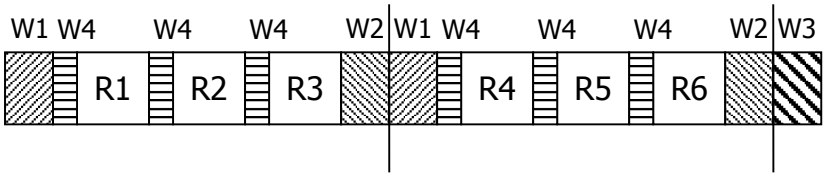
$W3$ = حافظه هرز ناشی از جا نگرفتن بلاک سوم در شیار

$$BF = \left\lfloor \frac{B}{R} \right\rfloor$$

حافظه هرز به ازاء هر بلاک $WB = W1 + W2 + \frac{W3}{TF}$

حافظه هرز به ازاء هر رکورد $WR = \frac{WB}{BF}$

روش دوم:

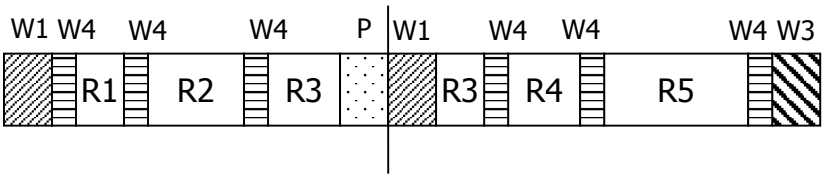


$P = W4$ = فضای هرز فیلدی که طول رکورد نوشته می شود

R = طول متوسط رکوردها

$$WR = P + \frac{1}{BF} \times \left(G + \frac{R}{2} + \frac{W3}{TF} \right)$$

روش سوم:



$$BF = \frac{B - P}{R + W4}$$

$$WB = W1 + P + BF \times W4 + \frac{W3}{TF}$$

$$WR = P + \frac{1}{BF} \times \left(G + P + \frac{W3}{TF} \right)$$

معیار سنجش یک نرم افزار: پیچیدگی خود نرم افزار، میزان حافظه‌ای که اشغال می کند، سرعت، کارایی بالا.

زمان حیات یک فایل:

از وقتی که یک فایل ایجاد می شود تا وقتی که پاک می گردد.

تغییر طول فایل به دو علت می باشد: ۱ــ تعداد رکوردهای فایل عوض شود ۲ــ طول رکوردهای فایل عوض شود.

مزایای بلاک بندی:

کاهش دفعات خواندن و نوشتن، کاهش حافظه هرز ناشی از وجود گپ بین رکوردها.

معایب بلاک بندی:

کار نرم افزاری بیشتر، مصرف حافظه ی اصلی بیشتر، بالا رفتن احتمال خطا در اطلاعات به علت افزایش میزان اطلاعات انتقالی در یک عمل ورودی و خروجی (هر چه حجم اطلاعات بالا رود حجم خطا هم بالا میرود)

باکت بندی:

مجموعه ای از تعدادی بلاک با حداقل طول یک بلاک می باشـــد. مزایا و معایب باکت بندی همان مزایا و معایب بلاک بندی است.

اگر بلاک بندی دا شته با شیم واحد خواندن می شود بلاک و اگر باکت بندی دا شته با شیم واحد خواندن و نو شتن می شود باکت و هر وقت با کل فایل کار دا شتیم سیستم عامل مطرح ا ست و هر وقت با رکورد کار داشتیم نرم افزار مطرح می شود.

اگر رکوردی دوپاره گردد در روش سوم بلاک بندی دا شتیم، دوباره باید به سراغ دیسک برویم؛ یعنی دوبار خواندن و نوشتن صورت می گیرد ولی در روش باکت بندی این مشکل را نداریم.

چگالی Load اولیه در فایل ها:

تعریف Load اولیه در فایل ها:

ایجاد اولیه ی یک فایل و اطلاعات اولیه ای که در فایل قرار می گیرد و لحظه Load اولیه فایل یک مقداری از بلاک ها را خالی می گذارند برای اضـــافه کردن اطلاعات بعدی. اگر پشـــت ســـر هم رکوردها را بنویسیم و یکدفعه فضای خالی بگذاریم، اگر رکوردی را اضافه بنماییم، همه رکوردها باید یک شیفت به جلو بخورند و این نرم‌افزاری است و بافری که تعیین می کنیم مقداری از آن را خالی می گذاریم و به سراغ بعدی می رویم.

$$\text{چگالی Load اولیه} = \frac{\text{مقدار فضای پر شده}}{\text{اندازه بلاک}} = \frac{LD}{B} < 1$$

مزایای چگالی Load اولیه:

۱ـــ حافظه ی Locality (همسایگی) بیشتر در فایل (در اثر ناحیه ی رزرو شده) همسایگی و همجواری و نزدیکی بیشتر، نظم منطقی فایل با نظم فیزیکی آن.

۲ـ تسهیل عملیات روی فایل. مثلاً درج رکورد یا جستجوی رکورد.

معایب چگالی Load اولیه:

افزایش حافظه ی هرز در Load اولیه (در اثر وجود ناحیه ی رزرو) و ســـبب افزایش طول خطی فایل می شود و از این رهگذر خواندن تمام فایل زمان گیرتر می شود.

همسایگی یا Locality:

میزان نزدیکی رکورد منطقاً بعدی در محیط فیزیکی به رکورد فعلی.

هرچه میزان همسایگی کم شود، زمان دسترسی به رکورد بعدی زیاد می شود.

سطوح نشانی دهی به فایل (آدرس دهی به فایل):

بســـته به این که از چه دیدگاهی به فایل نگاه می کنیم، میزان آدرس دهی فرق می کند و بحث بر روی سیستم فایل است.

سیستم فایل از دو قسمت تشکیل شده است:

۱ـ بخش منطقی

۲ـ بخش فیزیکی

بخش منطقی سیستم فایل:

وظیفه اش اجرای درخواست های کاربر است و یک سری عملیاتی را که کاربر احتیاج دارد، انجام می دهد و معمولاً عملیات عبارتند از: بازکردن فایل، خواندن فایل (رکورد)، و نوشـــتن فایل (رکورد)، بســـتن فایل و جستجوی یک رکورد.

بخش فیزیکی سیستم فایل:

وظیفه اش دریافت دســـتورات از بخش منطقی و تبدیل آن ها به فرامین کنترل کننده ی رســـانه ی ذخیره سازی است.

سطوح برخورد با فایل:

۱ـ سطح برخورد با کاربر

۲ـ سطح برخورد بخش منطقی سیستم فایل

۳ـ سطح برخورد بخش فیزیکی

سیستم فایل.

نشانی دهی کاربر (پردازشگر فایل):

منظور آدرس دهی به یک رکورد است؛ یعنی مشخص کردن یک رکورد خاص.

نوع نشانی دهی کاربر سه قسمت است:

۱ـــ نشانی دهی محتوایی: از نظر مشخص کردن محتوای یکی از فیلدها، و معمولاً این فیلد، کلید آن رکورد است.

۲ ـ نشانی دهی نسبی در یک فایل: یک شماره به رکورد می دهیم و وقتی کاربر رکورد مشخصی را صدا می کند آن شماره را می دهد.

۳ـــ نشانی دهی نمادی یا سمبلی: برای بعضی رکوردها اسم سمبلیک می گذاریم. مثلاً برای دانشجویانی که معدلشـــان بالاســـت واژه ی First یعنی رتبه ی اول و Second یعنی دانشـــجویان رتبه ی دوم به عنوان سمبلیک که صدا کردن آن رکورد از طریق این اسم می باشد.

آدرس دهی فیزیکی شامل این موارد می شود:

۱ـ شماره رسانه

۲ـ شماره استوانه

۳ـ شماره شیار

۴ـ شماره بلاک

دسترسی اطلاعات در یک فایل، اجرای درخواست کاربر توسط سیستم فایل:

۱ـ بازکردن فایل

۲ـ خواندن یک رکورد

۳ـ نوشتن یک رکورد

۴ـ بستن فایل

منظور از بازکردن یک فایل:

ابتدا فایل را روی رسانه پیدا کنیم و بعد مشخصات فایل از روی دیسک باید خوانده شود و سپس بررسی حق دستیابی کاربر

تخصیص بافر به فایل:

وقتی که می خواهیم با فایلی کار کنیم، خواندن و نوشتن از طریق بافر انجام می گیرد.

مشخصات فایل و بررسی حق دستیابی کاربر را راهنمای فایل می گویند.

راهنمای فایل:

۱ـ اسم فایل

۲ـ اسم صاحب فایل

۳ـ تعداد رکوردهای منطقی در فایل

۴ـ طول رکوردها (رکورد ثابت)

۵ـ آدرس اولین رکورد

۶ـ اندازه ی فایل

۷ـ تاریخ ایجاد فایل

۸ـ حقوق دستیابی به فایل (Password)

۹ـ تاریخ آخرین تغییرات در فایل