

جلسه هشتم

سرریزی:

بررسی مسئله ی سرریزی: به این چند سؤال ابتدا باید پاسخ داد.

۱- فضای لازم برای درج رکورد سرریزی چگونه انتخاب می شود؟

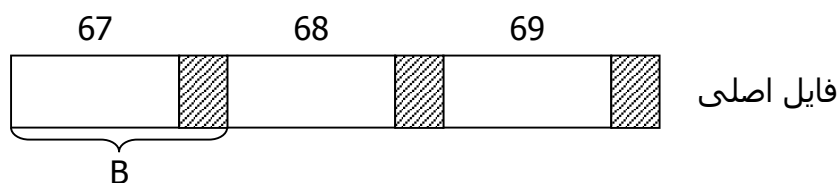
۲- فضای انتخاب شده چگونه در محیط فیزیکی (دیسک) به فایل تخصیص می یابد؟

۳- عمل درج با چه تکنیکی انجام می شود؟

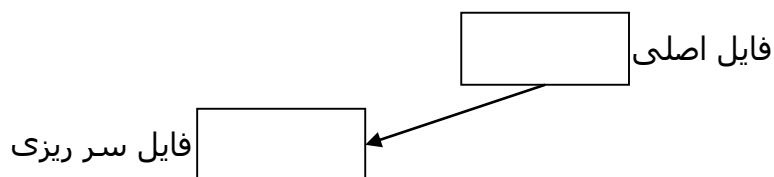
برای حالت ۱ سه راه حل پیشنهاد شده:

الف: در نظر گرفتن جا در هر بلاک
 ب: در نظر گرفتن فایل جداگانه
 ج: در نظر گرفتن جا در همان فایل اصلی

الف) موقعی که فایل را می سازیم در هر بلاک فایل اصلی یک مقداری فضای خالی در نظر می گیریم. هر چند که به نظر می رسد از نظر قوی بودن لوکالیتی رکوردها، راه خوبی باشد ولی عیب اش این است که پیش بینی چه مقدار جا را نداریم.



ب) در روش ب یک فایل اصلی را در نظر می گیریم و یک فایل سرریزی که سرریزی ها را داخل آن بریزیم و اشاره گر داریم از فایل اصلی به فایل سرریزی که این مشکل است از یک فایل به فایل دیگر اشاره کردن و زمان گیر است و تلف زمانی دارد.

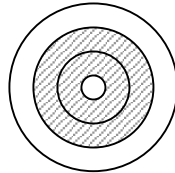


ج) بهترین روش و رایج ترین روش است. در همان فایل اصلی جایی برای سرریزی را در نظر می گیریم و لی باید ببینیم کجای فایل اصلی و دو روش وجود دارد.

الف: استوانه های جداگانه در نظر بگیریم
 ب: شیارهایی از هر استوانه برای سر ریزی همان استوانه در نظر گرفته شود (در انتهای هر استوانه)

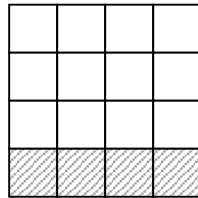
۲- تقسیم بندی سخت افزاری است:

الف) این راه حل مناسب نیست زیرا سبب می شود که لوکالیتی رکوردهای سر ریزی ضعیف شود و در نتیجه متوسط زمان استوانه جویی افزایش یابد.

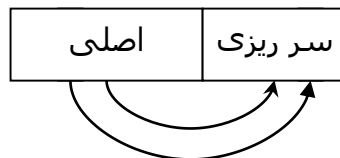


ب) این راه حل متوسط زمان استوانه جویی را کاهش می دهد زیرا رکوردهای سر ریزی هر استوانه در همان استوانه جای دارند. البته وقتی که ناحیه ی سر ریزی یک استوانه پر شود ناحیه ی دیگری برای درج سر ریزی ها باید ایجاد کرد (ناحیه سر ریزی اولیه و ثانویه) و یا این که فایل را سازمان دهی مجدد کرد.

استوانه ها



و به صورت نرم افزاری بهترین کار این است که ناحیه ی سر ریزی را بگذاریم در ناحیه ی فایل اصلی و لوکالیتی اش کم تر است.



الف: درج در اولین بلاک جادار در ناحیه سر ریزی
 ب: درج با Push Trough

۳- عمل درج در ناحیه ی سر ریزی با چه تکنیکی انجام می شود؟

الف) در این تکنیک، رکورد جدید مستقیماً وارد بلاکی از ناحیه ی سر ریزی می شود و در اولین مکان آزاد جای می گیرد (در اولین بلاک جادار). سپس از رکورد منطقی پیشین به رکورد درج شده نشانه رو ایجاد می شود و به ترتیب زنجیره ی رکوردهای

سر ریزی پدید می آید. در این تکنیک برای هر رکورد از ناحیه ی اصلی و ناحیه ی سر ریزی یک نشانه رو وجود دارد. البته ممکن است در بعضی از رکوردها محتوای این فیلد Null باشد.

در فیلد نشانه رو آخرین رکورد هر زنجیره، Null داریم به معنای این که پایان زنجیره است. ممکن است از یک بلاک چند اشاره گر به ناحیه ی سر ریزی داشته باشیم. روش درج ساده است ولی جست و جو وقت گیر می باشد.

ب) در روش ب درج مشکل است ولی جست و جو ساده تر است و اساس بر این دو نکته ی زیر می باشد:

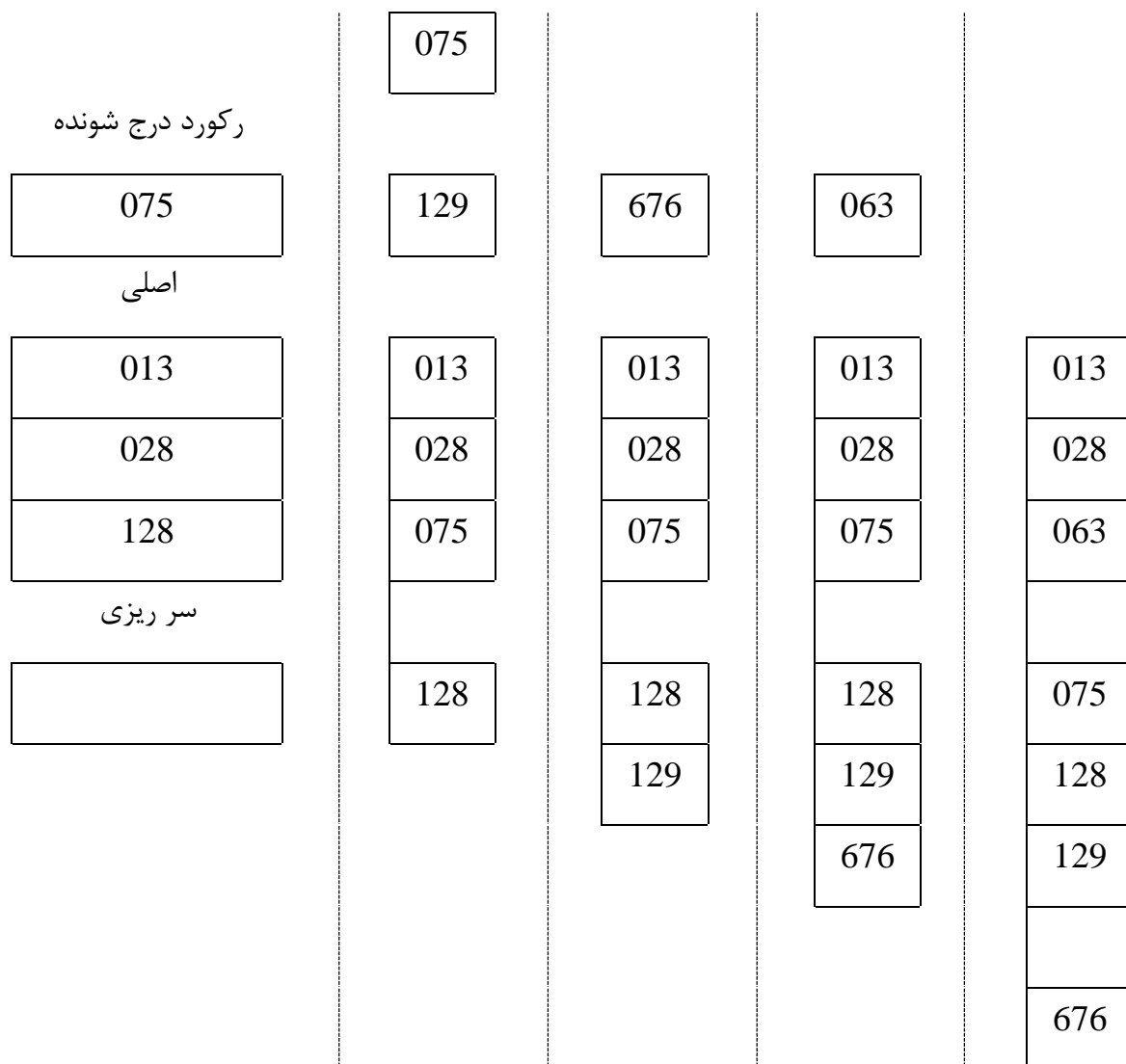
۱- نظم درون بلاکی همیشه وجود داشته باشد، حتی برای ناحیه ی سر ریزی.

۲- از هر بلاک ناحیه ی اصلی تنها یک اشاره گر به ناحیه سر ریزی اشاره کند.

وقتی یک رکورد درج می شود باید محل منطقی اش را پیدا کنیم در داخل بلاک و رکورد درج شده را در آن محل قرار دهیم و رکوردهای از آن به بعد را یک رکورد شیفت بدهیم جلو و ته فایل یک بلاک اضافه پیدا می کنیم و آن را می آوریم در ناحیه ی سر ریزی و کار تا آن جا پیدا می کند که دیگر شیفت کردن لازم نداشته باشیم.

در ناحیه ی اصلی داریم در لحظه ی شروع کار:

حال می خواهیم یک سری رکورد درج کنیم و رکورد درج شونده است مثلا



موارد استفاده ی ساختار:

این ساختار در محیط هایی به کار می رود که در آن ها نیاز به پردازش سریال فایل روی یکی از صفات خاصه کلید (مطرح) بوده. به علاوه واکنشی تک رکوردها از طریق مقدار کلید آن ها عمل رایجی باشد، در اغلب سیستم های تجاری — مدیریتی، این ساختار مورد استفاده قرار می گیرد.

ساختار فایل چند شاخصی Multi Indexed:

اهمیت این ساختار به خاطر این است که ۷۰٪ الی ۸۰٪ ساختار بانک های اطلاعاتی با این ساختار است.

در فایل های ترتیبی یک فایل اصلی داشتیم که مرتب بود و یک فایل جانبی داشتیم برای درج کردن و مشکل این بود که هیچ ارتباط منطقی بین آن ها نبود و برای رفع این اشکال فایل ترتیبی شاخص دار داشتیم و از یک سیستم شاخص استفاده می کردیم که کمک می کرد به دستیابی سریع.

سه ایراد عمده داشت:

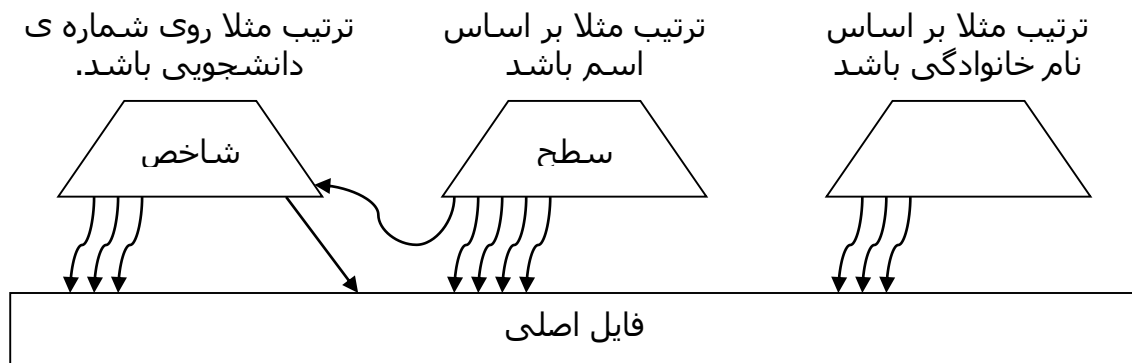
۱- ایستا بودن شاخصها بود.

۲- عدم تقارن بود.

۳- مسئله سر ریزی بود.

و برای رفع این معایب:

در فایل های چند شاخصی روش کار به این صورت است که یک فایل داریم به اسم فایل غیر اصلی و غیر ترتیبی است و حتی می تواند یک فایل پایل باشد.



فایل اصلی، غیر ترتیبی یا پایل میتواند باشد

شاخص متراکم، شاخص به همه ی رکوردها داریم.

شاخص دینامیک است، شاخص همراه تغییر رکوردها تغییر می کند.

این ساختار چنان است که پدیده ی عدم تقارن در آن وجود ندارد. زیرا روی تعدادی، حتی تمام صفات خاصه می توان شاخص داشت و مسئله ی رکوردهای سر ریزی به صورتی که در ساختار سوم مطرح بود در این جا وجود ندارد. یعنی درج رکوردهای جدید آسان تر و پویاتر است و بالاخره خود ساختار شاخص وضعیت پویا دارد و هم روند با تغییرات فایل داده ای، قابل تنظیم و به هنگام در آوردن است. اگر a تعداد صفات خاصه در فایل باشد، حداکثر a فایل شاخص می توان داشت.

از آن جا که در یک رکورد به طور متوسط، 'a' تا صفت خاصه وجود دارد، لذا به یک رکورد 'a' ساختار شاخص ناظر است. پس این ساختار در اساس از نظر فایل داده ای همان پایل است اما مجهز به یک سری استراتژی دستیابی قوی، پویا و سریع. کاربر می تواند هر تعداد از صفات خاصه ای که در فایل دارد، درخواست ایجاد شاخص کند و برای واکنشی سریع تک رکوردها، الزامی ندارد که حتما از کلید اصلی به عنوان آرگومان جست و جو استفاده نماید.

ضابطه ی انتخاب صفات خاصه ی شاخص: لزومی ندارد که روی تمام صفات خاصه، شاخص ایجاد نمود، می توان بین صفات خاصه قائل به اولویت شد و آن صفاتی را برگزید که در بیشترین درخواست ها به عنوان آرگومان جست و جو به کار برده می شوند.

اجزاء ساختار ترتیبی شاخص دار عبارتند از: ناحیه اصلی، ناحیه سر ریزی، مجموعه شاخص ها.

در فایل چند شاخصی داریم: هرچه تعداد صفات خاصه شاخص بیشتر باشد، عمل بازیابی کارآتر است و هرچه تعداد صفات خاصه شاخص بیشتر باشد، عدم تقارن کمتر است و ساختار فایل داده ای اصلی می تواند پایل باشد و ایجاد شاخص روی ترکیبات مختلف صفات خاصه امکان پذیر است.

فایل وارون، فایل است که روی تمام فیلدهای آن شاخص داشته باشیم.

در فایل چند شاخصی می توان بین صفات خاصه اولویت قائل شد و برای ایجاد شاخص آن فیلدهایی را انتخاب کرد که در بیشتری پرس و جو ها به کار برده می شوند.

تعداد مدخل های شاخص در سطح اول برای شاخص های مختلف ممکن است یکسان نباشد.

فایل های شاخص تأمین کننده ی استراتژی دستیابی برای فایل داده ای هستند.

در فایل داده ای مقدار فیلدی می تواند Null یا ناشناخته باشد.

شاخص گذاری جزو نوع دستیابی تصادفی است.

فایلی شامل اطلاعات ثبت نام دروس دانشجویان با ۲۰۰۰ رکورد داریم. هر دانشجو به طور متوسط ۴/۵ درس اخذ کرده است.

اگر روی فیلد درس اخذ شده (که تکرار شونده است) شاخص ایجاد کنیم تعداد مدخل های این شاخص چقدر خواهد بود؟

$$2000 \times 4.5 = 9000$$

تفاوت بین شاخص اولیه و ثانویه آن است که در شاخص اولیه کلید تکراری وجود ندارد ولی در شاخص ثانویه ممکن است کلید تکراری داشته باشیم.

اگر در فایلی که شاخص اولیه و ثانویه دارد بر اثر اصلاح، کلید اولیه تغییر کند، چه تنظیم دیگری را باید انجام دهیم؟

ممکن است لازم باشد تا شاخص اولیه و تمام شاخص های ثانویه تغییر کنند.
بر روی فایل های ترتیبی و پایل و فایل پایل مرتب شده می توان شاخص ایجاد کرد.