

INDEXING

Basic Index Concepts

2

□ ایندکسها باعث افزایش سرعت دستیابی به اطلاعات در جدول می شوند.

□ card catalog in a library (author, title, subject)

□ به مجموعه ای از یک یا چند خصیصه که برای جستجوی رکوردها در جدول استفاده می شوند، کلید جستجو گفته می شود.

□ در ساده ترین شکل، فایل ایندکس شامل رکوردهایی به شکل زیر است:

search-key	pointer
------------	---------

□ به چنین رکوردی، ورودی ایندکس گفته می شود.

Basic Index Concepts, Cont.

3

- فایل ایندکس یک فایل است و بسیاری از مشکلات فایل‌های داده را دارد.
- فایل‌های ایندکس معمولاً بسیار کوچکتر از فایل اصلی هستند.
 - 10% - 25%
- دو نوع اصلی از ایندکسها وجود دارد:
 - ایندکسهای مرتب: ورودی‌های ایندکس به صورت مرتب براساس کلید جستجو ذخیره می‌شوند.
 - ایندکس‌های hash: کلید جستجو با استفاده از یک "تابع hash" به صورت یکنواخت در باکتهای توزیع شده است.

Index Evaluation Metrics

4

متریکهای ارزیابی ایندکس

- زمان درج، حذف و بروزرسانی در پردازش تراکنش برخط
- زمان دسترسی در سیستمهای تصمیم گیری برای:
- پرس و جوهای نقطه ای: رکوردهایی با مقدار خاص یک خصیصه
- پرس و جوهای دامنه: رکوردهایی با مقدار خصیصه در دامنه معین
- سربار حافظه

Ordered Indices

5


- ایندکسی که کلید جستجوی آن، بر اساس ترتیب اطلاعات فایل داده باشد ایندکس اولیه نامیده می شود.
- ایندکس خوشه بندی هم گفته می شود.
- کلید جستجوی ایندکس اولیه معمولاً کلید اصلی است.
- ایندکسی که کلید جستجوی آن بر مبنای ترتیب فایل داده نباشد، ایندکس ثانویه گفته می شود.
- همچنین ایندکس غیر خوشه بندی هم گفته می شود.

Dense Index Files

6

□ به ایندکسی که برای هر مقدار کلید جستجو در فایل داده شامل رکورد ایندکس باشد، ایندکس چگال گفته می شود.

10101	→	10101	Srinivasan	Comp. Sci.	65000	↙
12121	→	12121	Wu	Finance	90000	↘
15151	→	15151	Mozart	Music	40000	↙
22222	→	22222	Einstein	Physics	95000	↘
32343	→	32343	El Said	History	60000	↙
33456	→	33456	Gold	Physics	87000	↘
45565	→	45565	Katz	Comp. Sci.	75000	↙
58583	→	58583	Califieri	History	62000	↘
76543	→	76543	Singh	Finance	80000	↙
76766	→	76766	Crick	Biology	72000	↘
83821	→	83821	Brandt	Comp. Sci.	92000	↙
98345	→	98345	Kim	Elec. Eng.	80000	↘



Dense Index Files, Cont.

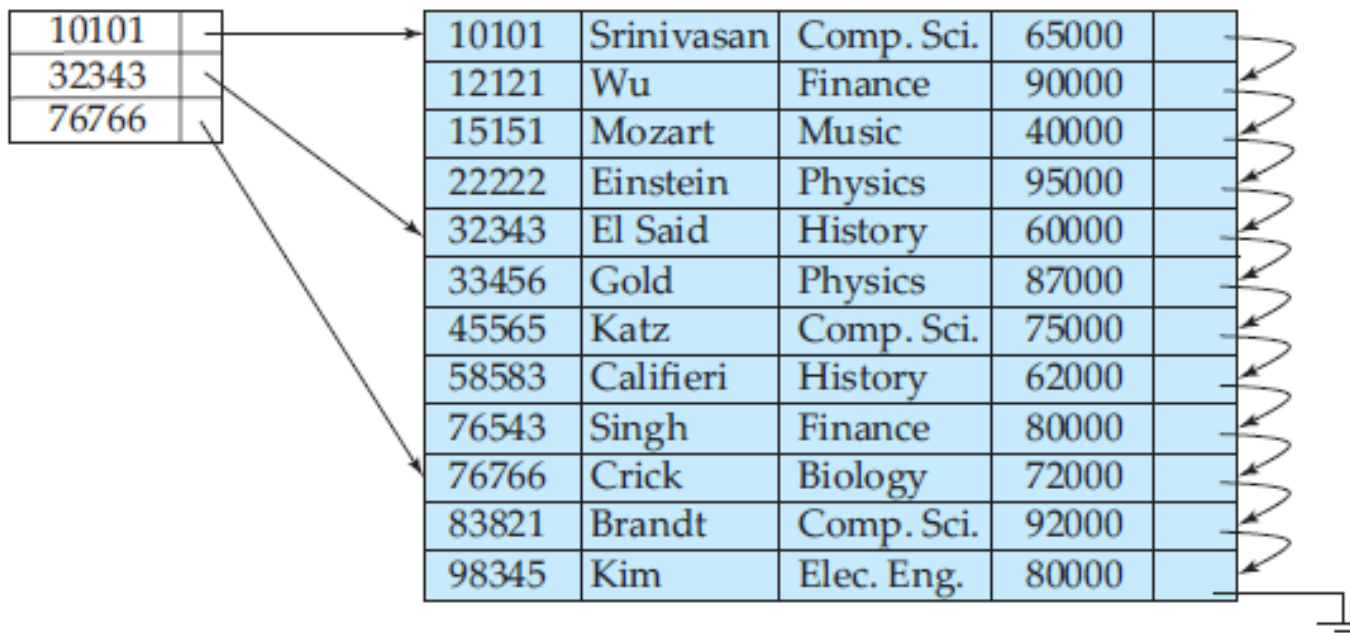
7

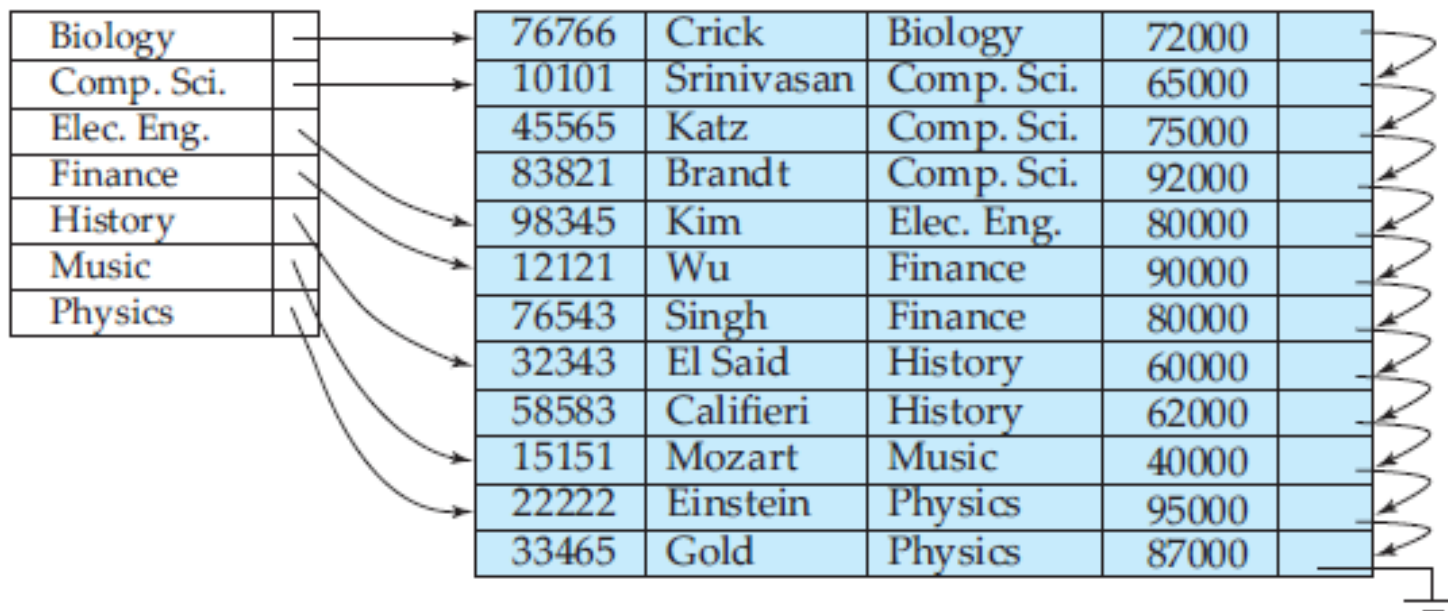
- برای یافتن رکورد(ها)ی که مقدار کلید جستجوی آن K است:
 - رکورد ایندکس با مقدار کلید جستجوی K را پیدا کن.
 - اشاره گر از رکورد ایندکس به رکورد(ها)ی داده را دنبال کن.
- برای حذف یک رکورد:
 - مکان رکورد را در فایل داده تعیین کن.
 - رکورد را از فایل داده حذف کن.
 - اگر رکورد حذف شده تنها رکورد با مقدار کلید جستجوی موردنظر باشد، کلید جستجو از ایندکس حذف میشود.
- برای درج رکورد:
 - بااستفاده از مقدار کلید جستجو، در ایندکس جستجو کن.
 - اگر مقدار کلید جستجو در ایندکس وجود ندارد، آن را درج کن.
 - رکورد را در فایل داده درج کن. (در مکان صحیح)
 - اشاره گری از ورودی ایندکس به رکورد داده ایجاد کن.

Sparse Index Files

8

- ایندکسی که فقط برای برخی مقادیر کلید جستجو شمل رکوردهای ایندکس باشد، ایندکس اسپارس است.
- معمولاً برای هر بلاک فایل داده یک ورودی ایندکس وجود دارد.





Sparse Index Files, Cont.

10

- برای یافتن رکوردی با مقدار کلید جستجوی K:
 - رکورد ایندکس یا بزرگترین مقدار کلید جستجو کمتر یا مساوی K را پیداکن.
 - در فایل به صورت ترتیبی از رکوردی که رکورد ایندکس اشاره میکند جستجو کن.
- برای حذف رکورد:
 - محل رکورد در فایل داده را تعیین کن.
 - رکورد را حذف کن.
 - اگر رکورد حذف شده تنها رکورد با مقدار کلید جستجوی معین است و ورودی برای آن در ایندکس وجود دارد، ورودی ایندکس را با مقدار کلید جستجوی بعدی فایل داده جایگزین کن. اگر مقدار کلید بعدی قبلاً در ایندکس درج شده است، رکورد را از ایندکس حذف کن.
- برای درج رکورد: (اگر برای هر بلاک یک ورودی ایندکس وجود داشته باشد)
 - با استفاده از مقدار کلید جستجوی رکورد، در ایندکس جستجو کن.
 - اگر ورودی ایندکس به بلاکی اشاره میکند که فضای خالی دارد، رکورد را در جای مناسب در بلاک درج کن.
 - اگر ایندکس به بلاک پر اشاره میکند، یک بلاک جدید تخصیص بده و اولین مقدار کلیدی که در بلاک جدید دیده می شود را به ایندکس اضافه کن.

Sparse Index Files, Cont.

11

□ مزایا نسبت به ایندکسهای چگال:

- نیاز به فضای کمتری دارد.
- نگهداری کمتری برای درج و حذف نیاز دارد

□ معایب:

- جستجوی رکورد کندتر است به خصوص زمانی که برای یک ورودی ایندکس بیش از یک بلاک وجود داشته باشد.

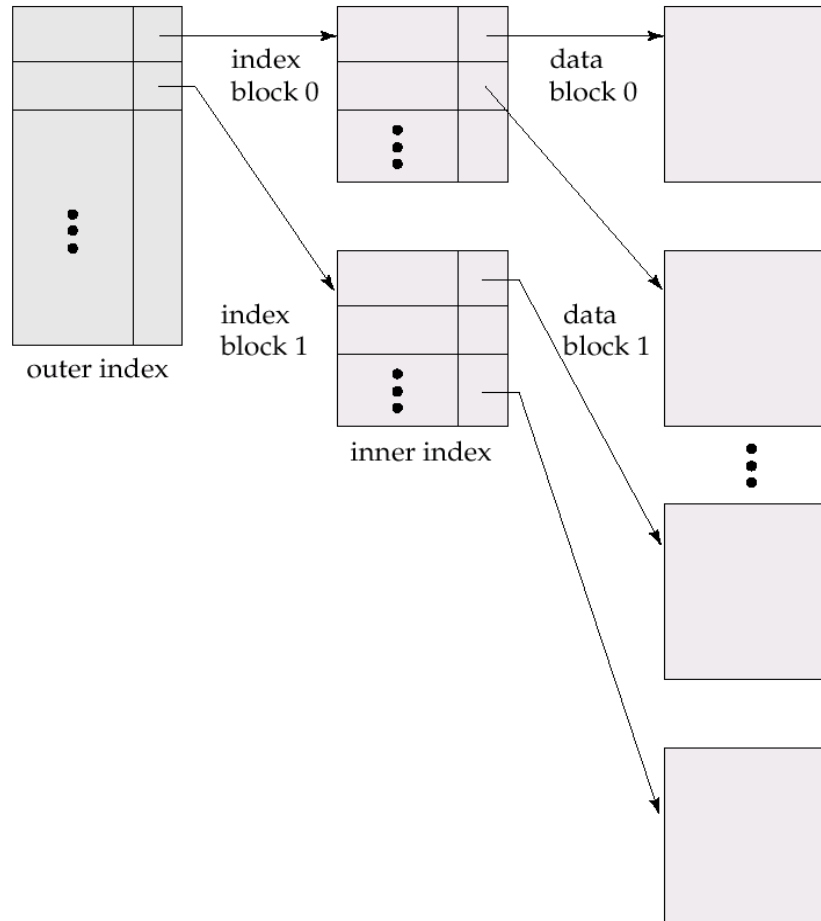
Multilevel Index

12

- یک روش برای افزایش کارایی این است که هکه بلاکهای ایندکس در حافظه نگهداری شوند ولی برخی اوقات ایندکس حجم زیادی دارد و این امکان وجود ندارد.
- در چنین حالتی، ایندکس را می توان به صورت یک فایل ترتیبی در دیسک در نظر گرفت و یک ایندکس اسپارس بر روی آن ایجاد کرد.
 - ایندکس خارجی: اسپارس
 - ایندکس داخلی: چگال یا اسپارس
- اگر ایندکس خارجی هنوز حجم زیادی برای نگهداری در حافظه داشته باشد، سطح دیگری از ایندکس ایجاد می شود.

Multilevel Index, Cont.

13



□ با هر عمل درج یا حذف ممکن است لازم باشد ایندکسهای کلیه سطوح بروزرسانی شوند.

Secondary Indices

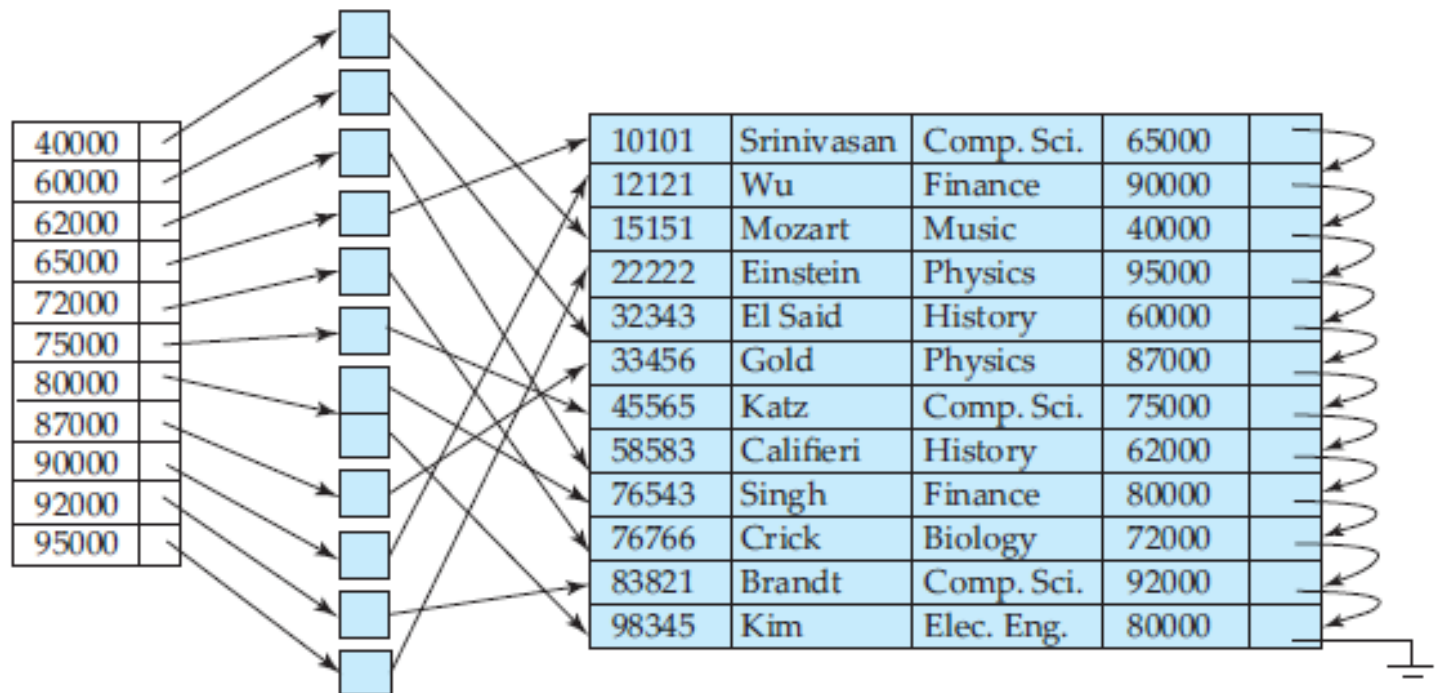
14

□ ایندکس ثانویه زمانی استفاده میشود که جدول با استفاده از کلیدی به جز آنچه مرتب شده است جستجو میشود.

- Suppose *account* is sorted by account number, and
- Searches are based on branch, or searching for a range of balances.

□ هر ورودی ایندکس اشاره دارد به:

- رکورد واحدی با مقدار کلید جستجوی معین (کلید کاندیدا)
- باکتهی که شامل پوینترهایی به همه رکوردهایی است که مقدار کلید جستجوی معینی را دارند. (کلید غیر کاندیدا)



Index Performance

16

- اگرچه ایندکسها کارایی را افزایش میدهند ولی مشکلات زیر را هم دارند:
 - همه ایندکسها باید در صورت درج یا حذف بروزرسانی شوند.
 - با افزایش حجم فایل ایندکس، کارایی کاهش می یابد، زیرا بلاکهای سرریز زیادی ایجاد می شوند و بنابراین سازماندهی مجدد ایندکس (حذف و بازسازی) لازم است.
 - پیمایش ترتیبی فایل با ترتیب ثانویه ممکن است پرهزینه باشد. در بدترین حالت هردسترسی رکورد ممکن است بلاک جدیدی از دیسک را بخواند.
- بنابراین تعداد رکوردهایی که بازیابی می شوند یک حد بالا برای تعداد بلاک های داده ای است که در زمان پویش فایل در پرس و جوهای دامنه بازیابی میشوند.

create index <index-name> **on** <relation-name> (<attribute-list>);



create index *studentID_index* **on** *student*(ID);

create index *dept_index* **on** *instructor* (*dept_name*);

create index *inst_sal_idx* **on** *instructor* (*salary*);