



دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

# طراحی پایگاه داده‌ها

( فصل پنجم : معماری سه سطحی پایگاه داده‌ها )

مهدی دادبخش

[mahdi.dadbakhsh@sharif.edu](mailto:mahdi.dadbakhsh@sharif.edu)

شماره درس : ۴۰۳۸۴

یکشنبه - سه‌شنبه ( ۱۶:۳۰ الی ۱۸:۰۰ )

۱۴۰۱ - ۱۴۰۲

معماری سه سطحی ANSI

اجزای معماری سه سطحی

سطح (دید) خارجی

سطح (دید) ادراکی

سطح (دید) داخلی

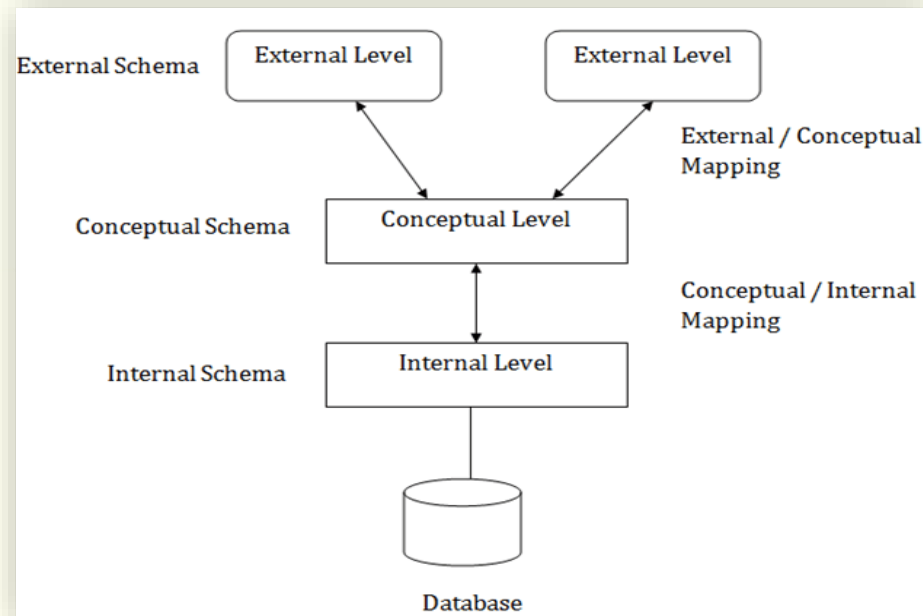
نکاشت بین سطوح

شاخص

پایان

## معماری سه سطحی ANSI

- این معماری به عنوان استاندارد برای طراحی سیستم‌های پایگاه داده به کار می‌رود.
- ANSI / SPARC مخفف American National Standards Institute, Standards Planning and Requirement Committee است.
- این استاندارد اولین بار در سال ۱۹۷۵ برای طراحی سیستم‌های مدیریت پایگاه داده پیشنهاد شد.
- هدف اصلی معماری سه سطحی این است که چندین کاربر را قادر می‌سازد تا با یک دید شخصی به داده‌های مشابه دسترسی داشته باشند، درحالی‌که داده‌های اساسی را تنها یک بار ذخیره می‌کنند. بنابراین دید کاربر را از ساختار فیزیکی پایگاه داده جدا می‌کند. این تفکیک به دلایل زیر مطلوب است:
  - کاربران مختلف به دیدهای متفاوتی از داده‌ها نیاز دارند.
  - کاربران نباید نگران پیاده سازی فیزیکی و عملکرد داخلی پایگاه داده ( نظیر تکنیک‌های فشرده سازی، رمزنگاری داده‌ها و غیره ) باشند.
  - DBA باید بتواند ساختار مفهومی پایگاه داده را بدون تاثیرگذاری روی کاربر تغییر دهد.
  - ساختار داخلی پایگاه داده نیز باید تحت تاثیر تغییرات در جنبه‌های فیزیکی ذخیره سازی قرار نگیرد.



## اجزای معماری سه سطحی

- ۱- کاربر User
- ۲- زبان میزبان ( HL - Host Language ) : برنامه نویسان، برنامه های کاربردی را با استفاده از یک زبان های سطح بالا نظیر Visual basic، Java و Visual C پیاده سازی می کنند. زبان سطح بالائی که علاوه بر داشتن امکانات گوناگون دارای دستوراتی برای تعریف و کار با داده هستند، زبان میزبان نامیده می شوند.
- ۳- زبان داده ای فرعی ( DSL – Data Sub Language ) : زیرمجموعه ای از زبان میزبان که مختص عملیات ذخیره و بازیابی اطلاعات از پایگاه داده است زبان فرعی داده نام دارد و ترکیبی از سه زبان DDL ، DML و DCL است.
- ۴- سطح یا دید خارجی
- ۵- سطح یا دید ادراکی
- ۶- سطح یا دید داخلی
- ۷- فایل های فیزیکی
- ۸- سیستم مدیریت پایگاه داده ها
- ۹- مدیر پایگاه داده ها ( DBA ) : فردی است که مسئول کنترل عملیات کل سیستم پایگاه داده است و کلیه فعالیت های سیستم پایگاه داده را هماهنگ می کنند. این فرد باید درک خوبی از منابع و نیازهای اطلاعاتی کل سازمان داشته باشد و برای حصول اطمینان از اینکه داده موردنیاز قابل دسترس کاربران قرار می گیرد با آنها در ارتباط باشد. بعضی از وظایف DBA شامل : تعریف شماها توسط DDL ، تعریف ساختار ذخیره سازی و متدهای دسترسی توسط DDL ، اصلاح شما و سازماندهی فیزیکی ، اعطای مجوز دسترسی پایگاه داده به کاربران ، تعیین قیدهای جامعیت ، عامل ارتباطی کاربران ، نظارت اجرا و واکنش برای تغییر در صورت نیاز و برقراری دیکشنری داده .



## سطح ( دید ) خارجی

- سطح خارجی در بالاترین سطح در معماری سه سطحی و نزدیکترین سطح به کاربر است و همچنین به عنوان "دید" نیز شناخته می‌شود.
- سطح خارجی فقط محتوای پایگاه داده مربوطه را به صورت view به کاربران نشان می‌دهد و بقیه داده‌ها را پنهان می‌کند.
- بنابراین کاربران مختلف می‌توانند پایگاه داده را به صورت یک دید متفاوت بر اساس نیازهای فردی خود ببینند.
- چندین کاربر نیز ممکن است دید یکسانی داشته باشند.
- سطح خارجی، دید کاربر از داده‌های ذخیره شده در پایگاه داده است. منظور از دید کاربر ( user view ) قسمتی از پایگاه داده است که کاربر با آن سروکار دارد.
- دید هر کاربر باید تعریف شود. به تعریف و شرح دید کاربر شمای خارجی ( external schema ) می‌گویند.
- برای تعریف شمای خارجی از یک مدل داده استفاده می‌شود که معمولاً همان مدلی است که در سطح ادراکی بکار رفته است.
- این سطح، نزدیکترین سطح به کاربر است و با روش‌های مشاهده داده‌ها توسط کاربران مختلف سروکار دارد.
- این سطح شامل دستورات DML می‌باشد.

External  
View

Empno	Ename
-------	-------

Empno	Ename	Salary	DeptNo
-------	-------	--------	--------



## سطح ( دید ) ادراکی

- سطح ادراکی یا مفهومی در سطح پایین‌تری نسبت به سطح خارجی قرار دارد و به سطح منطقی نیز معروف است.
- سطح ادراکی ساختار کل پایگاه داده را توصیف می‌کند.
- سطح ادراکی توصیف می‌کند که چه داده‌هایی باید در پایگاه داده ذخیره شوند و همچنین چه رابطه‌ای بین آن داده‌ها وجود دارد.
- برنامه‌نویسان و مدیران پایگاه داده در این سطح کار می‌کنند.
- سطح ادراکی به نحوه ذخیره داده‌های موجود در پایگاه داده اهمیت نمی‌دهد.
- سطح ادراکی، دید طراح پایگاه داده از داده‌های ذخیره شده در پایگاه داده است. داده‌های دنیای واقعی آنطور که واقعاً هستند توسط طراح پایگاه داده مدل می‌شوند.
- برای تعریف سطح ادراکی از یک ساختار یا مدل داده استفاده می‌شود که شمای ادراکی ( conceptual schema ) نامیده می‌شود.
- برای کسب استقلال داده، شمای ادراکی تنها درگیر معنی داده است و جنبه‌های نمایش داده، سازماندهی فیزیکی و استراتژی‌های دستیابی ندیده گرفته می‌شود.
- این سطح شامل دستورات DDL و DCL است و دستورات DML را شامل نمی‌شود.

Global view

### EMPLOYEE

Empno : Integer(4) Key  
Ename : String(15)  
Salary : String (8)  
Deptno : Integer(4)  
Post : String (15)



## سطح (دید) داخلی

- این پایین ترین سطح در معماری سه سطحی است و به آن سطح فیزیکی نیز گفته می شود.
- سطح داخلی چگونگی ذخیره داده ها در پایگاه داده را توصیف می کند. در پایین ترین سطح این داده ها در حافظه های جانبی ( external hard disk ) به صورت بیت و در سطح کمی بالا می توان گفت که داده ها در فایل ها و پوشه ها ذخیره می شوند.
- سطح داخلی، دید طراح پایگاه داده از محیط فیزیکی ذخیره سازی و در واقع فایل های محیط فیزیکی است که توسط شمای داخلی ( internal schema ) توصیف می شود.
- شمای داخلی نحوه نمایش فیزیکی داده هایی را که در شمای ادراکی شرح داده شده را مشخص می کند.
- سطح داخلی به طور کلی شامل فعالیت های زیر می باشد :
  - تخصیص فضای ذخیره سازی
  - مسیرهای دسترسی ( به عنوان مثال : مشخص کردن کلیدهای اصلی و ثانویه، شاخص ها، اشاره گر ها و غیره )
  - تکنیک های فشرده سازی و رمز گذاری داده ها
  - بهینه سازی ساختارهای داخلی

Internal view

STORED\_EMPLOYEE record length 60

Empno : 4 decimal offset 0 unique  
Ename : String length 15 offset 4  
Salary : 8,2 decimal offset 19  
Deptno : 4 decimal offset 27  
Post : string length 15 offset 31



## نگاشت بین سطوح

- سه سطح معماری پایگاه داده مستقل از هم نیستند و باید بین آنها مطابقت و تناظری وجود داشته باشد، که به آن نگاشت بین سطوح می‌گویند.
- اساساً دو نوع نگاشت بین سطوح معماری پایگاه داده وجود دارد :
- **نگاشت ادراکی / داخلی ( C/I : Conceptual to Internal )**  
نگاشت ادراکی / داخلی بین سطوح ادراکی و داخلی قرار دارد و نقش آن تعریف مطابقت بین رکوردها و فیلدهای سطح ادراکی و فایل‌ها و ساختارهای داده سطح داخلی است.
- **نگاشت خارجی / ادراکی ( E/C : External to Conceptual ) :**  
نگاشت خارجی / ادراکی بین سطوح خارجی و ادراکی قرار دارد و نقش آن تعریف تناظر بین دیدگاه‌های خارجی و مفهومی است.





## شاخص (ایندکس)

- شاخص گذاری ( Indexing ) برای بهینه سازی عملکرد پایگاه داده با به حداقل رساندن تعداد دسترسی های دیسک هنگام پردازش یک پرس و جو استفاده می شود.
- ایندکس نوعی ساختار داده است که برای مکان یابی و دسترسی سریع به داده ها در جدول پایگاه داده استفاده می شود.
- می توان بر روی یک (یا چند) ستون از جدول، ایندکس تعریف کرد. وجود ایندکس بر روی مقادیر یک ستون باعث می شود به رکوردهای حاوی شرط **where** در فایل مربوط به یک جدول با سرعت بیشتری دسترسی داشته باشیم.
- **نکته :** در سیستم های جدولی، خود سیستم روی کلید اصلی ( PK ) ایندکس خودکار ( Automatic Index ) ایجاد می کند.

### ■ ساختار ایندکس :

فیلد اول ( Search key ) :

کلید جستجو است که حاوی یک کپی از کلید اصلی یا کلید کاندید جدول است. مقادیر کلید اصلی به ترتیب مرتب شده ذخیره می شوند تا بتوان به راحتی به داده های مربوطه دسترسی داشت.

فیلد دوم ( Data Reference ) :

شامل مجموعه ای از اشاره گر ها می باشد و آدرس بلوک هایی از دیسک را نگه می دارد که حاوی مقدار کلید مورد نظر است.

Search key	Data Reference
------------	----------------

### ■ ایجاد ایندکس :

**B-Tree Index :** CREATE INDEX *index\_name* ON *table\_name* ( *column\_name* ) [CLUSTERED]

**Hash Index :** CREATE INDEX *index\_name* ON *table\_name* USING HASH ( *column\_name* )

### ■ حذف ایندکس :

DROP INDEX *index\_name*

ایندکس Hash

ایندکس B-Tree

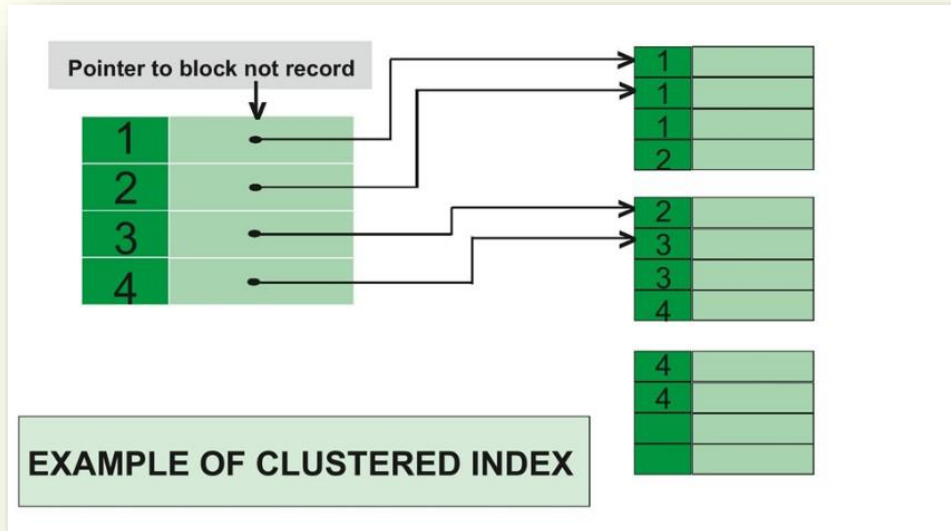
روش Non-clustered

روش Clustered



## روش خوشه‌ای (Clustered Index)

- در روش خوشه‌ای، همه‌ی داده‌های یک جدول با یک ترتیب مشخص به صورت فیزیکی در هارد دیسک ذخیره خواهند شد. از این رو یک جدول فقط یک ایندکس خوشه‌ای دارد که برای تمام داده‌ها اعمال می‌شود. برای توضیح مفهوم این روش از مثال ساده‌ی یک دفترچه تلفن استفاده می‌کنیم. در یک دفترچه تلفن شماره‌های افراد بر اساس ترتیب حروف الفبا قرار گرفته است و با پیدا کردن نام یک شخص به صورت مسقیم می‌توانید به شماره‌ی آن شخص دست پیدا کنید.
- این نوع ایندکس به صورت اتوماتیک بر روی همه جداول و بر اساس کلید اصلی بر روی پایگاه داده ایجاد می‌شود.
- در شکل زیر یک ایندکس خوشه‌ای و نحوه قرار گرفتن داده‌ها در آن مشاهده می‌شود:

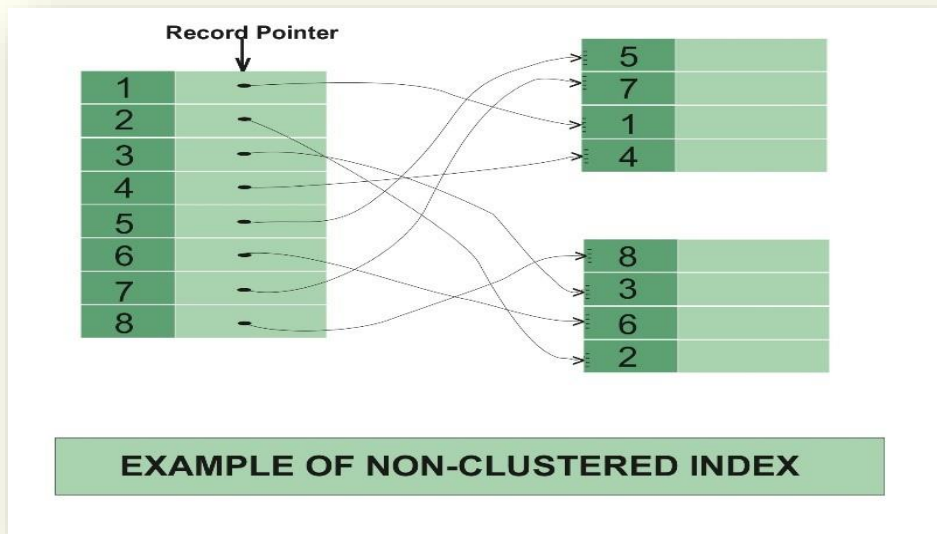


- **مزایای روش خوشه‌ای :**  
این روش برای جستجوهای که به صورت بازه‌ای هستند بسیار کارآمدتر است. مزیت دیگر این روش این است که زمانی که می‌خواهیم جستجو را به صورت مقایسه‌ای از ستون‌های بزرگتر یا کوچکتر از یک مقدار مشخص انجام دهیم کاربردی‌تر است مخصوصاً زمانی که تعداد داده‌ها زیاد است و داده‌ی تکراری نداریم.
- **معایب روش خوشه‌ای :**  
متأسفانه این روش برای ستون‌هایی که تغییر می‌کنند و به روزرسانی دارند مناسب نیست چرا که با هر تغییر تمام داده‌ها باید دوباره مرتب‌سازی شوند. همچنین در این حالت محدودیت ایندکس گذاری برای یک جدول داریم به این صورت که در یک جدول فقط یک نوع ایندکس گذاری وجود دارد که برای تمام داده‌ها اعمال می‌شود.

موضوع تحقیق : ایندکس خوشه‌ای را کامل شرح دهید.

## روش غیر خوشه‌ای ( Non-Clustered Index )

- این روش با ترتیب ذخیره‌ی داده‌ها کاری ندارد و تنها یک مقدار و pointer به داده‌ی مورد نظر اختصاص می‌دهد، مانند لیست واژه‌های مهم در انتهای یک کتاب و شماره صفحاتی که این واژه‌ها در آن کتاب به کار رفته است.
- این ایندکس‌ها را طراح پایگاه داده با توجه به نیاز می‌تواند ایجاد کند. البته توجه داشته باشید که ایجاد ایندکس غیرخوشه‌ای باید توسط یک متخصص انجام شود چرا که در صورت ایندکس گذاری اشتباه، دسترسی به اطلاعات کندتر انجام خواهد شد.
- در شکل زیر یک ایندکس غیرخوشه‌ای و نحوه قرار گرفتن داده‌ها در آن مشاهده می‌شود:



- مزایای روش خوشه‌ای :**  
از مزایای این روش می‌توان به سریع‌تر شدن روند جستجو در داده‌ها و مرتب سازی سریع‌تر نتایج اشاره کرد. همچنین در این حالت ما محدودیتی برای اعمال تعداد ایندکس در یک جدول نداریم و می‌توانیم از چندین ایندکس استفاده کنیم.  
این روش برای جستجوهای که مقایسه بین داده‌ها به دنبال تساوی است کارآمد می‌باشد.
- معایب روش خوشه‌ای :**  
این روش از روش خوشه‌ای کندتر است و برای جستجوهای که نتایج زیادی دارند کاربردی نیست.

موضوع تحقیق : ایندکس غیرخوشه‌ای را کامل شرح دهید.

# ایندکس B-Tree

- ایندکس B-Tree برای انواع شرطهای مقایسه‌ای و بازه‌ای کاربرد دارد.

- تعریف درخت B-Tree :**

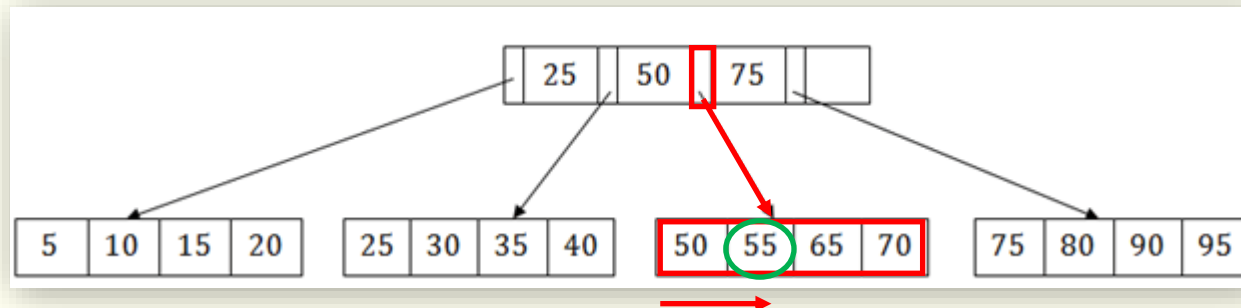
یک درخت جستجوی دودویی متعادل است و از یک فرمت اندیس گذاری چند سطحی پیروی می‌کند. در این درخت، گره‌های برگ اشاره‌گرهای داده واقعی را نشان می‌دهند و از طریق لیست پیوندی به هم مرتبط می‌شوند. بنابراین از دسترسی تصادفی و دسترسی متوالی پشتیبانی می‌کند. درخت B-Tree تضمین می‌کند که تمام برگ‌ها در یک سطح باشند.

- ساختار درخت B-Tree :**

در این درخت، هر گره برگ در فاصله مساوی از گره ریشه قرار دارد. این درخت شامل یک گره داخلی و گره برگ می‌باشد.

- جستجو در درخت B-Tree :**

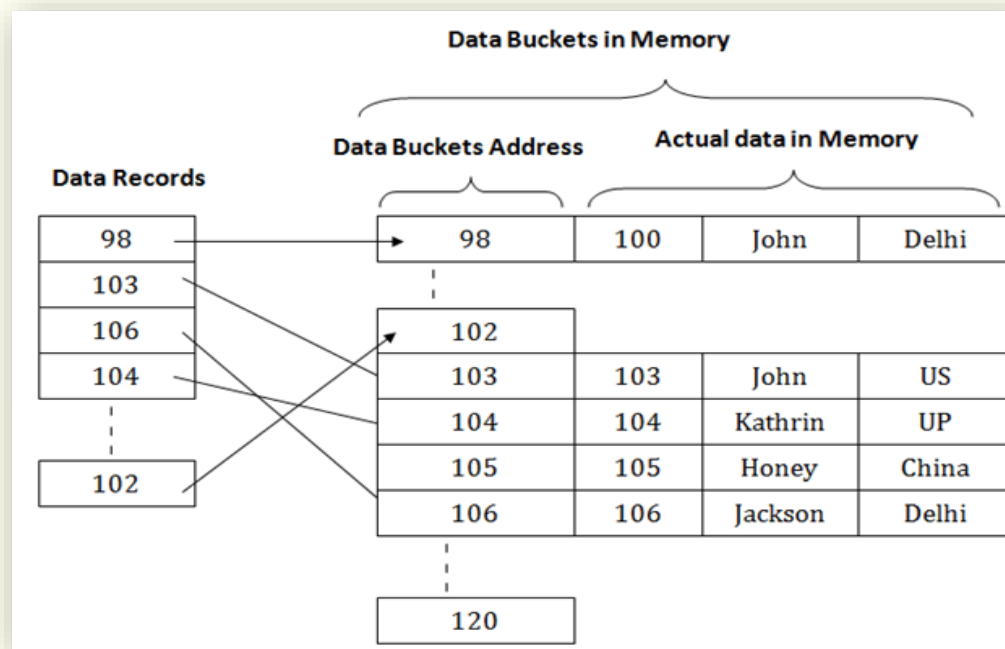
فرض کنید می‌خواهیم ۵۵ را در درخت B-Tree زیر جستجو کنیم. ابتدا گره واسطه را واکشی می‌کنیم که به گره برگ هدایت می‌شود که حاوی رکورد ۵۵ است. بنابراین در گره واسطه یک شاخه بین ۵۰ تا ۷۵ پیدا می‌کنیم. سپس در پایان به گره برگ سوم هدایت می‌شویم. در این حالت، DBMS یک جستجوی متوالی را برای یافتن ۵۵ انجام می‌دهد.



موضوع تحقیق : ایندکس B-Tree را کامل شرح دهید.

## ایندکس Hash

- ایندکس مبتنی بر درهم سازی ( Hash ) وقتی کارایی دارد که فقط شرط تساوی بر روی مقادیر ستون اندیس گذاری شده داشته باشیم.
- در یک ساختار پایگاه داده عظیم، جستجوی در تمام مقادیر ایندکس و رسیدن به داده‌های مورد نظر بسیار ناکارآمد است. تکنیک Hash برای محاسبه مکان مستقیم یک رکورد داده روی دیسک بدون استفاده از ساختار ایندکس استفاده می‌شود.
- در این تکنیک، داده‌ها در بلوک‌های داده‌ای که آدرس آنها با استفاده از تابع hash تولید می‌شود، ذخیره می‌گردد. مکان حافظه که این رکوردها در آن ذخیره می‌شوند به عنوان data bucket یا بلوک های داده شناخته می‌شود.
- یک تابع hash می‌تواند هر یک از مقادیر ستون را برای تولید آدرس انتخاب کند. البته بیشتر اوقات، از کلید اصلی برای تولید آدرس بلوک داده استفاده می‌کند.



موضوع تحقیق : ایندکس Hash را کامل شرح دهید.



## پایان فصل پنجم

مهدی دادبخش

*[mahdi.dadbakhsh@sharif.edu](mailto:mahdi.dadbakhsh@sharif.edu)*

۱۴۰۱ – ۱۴۰۲