

TRƯỜNG ĐÀO TẠO LẬP TRÌNH VIÊN VÀ QUẢN TRỊ MẠNG QUỐC TẾ BACHKHOA-APTECH

---

# Bài 11

## Indexes

# Tóm tắt

- Giới thiệu Index.
- Mô tả hiệu suất Index.
- Kiến trúc của Index.
- Tạo Index



# Giới thiệu

- Khi thực hiện một truy vấn, SQL Server 2014 sử dụng các **index (chỉ mục)** để tìm dữ liệu.
- Nếu không định nghĩa bất kỳ một index nào cho việc tìm kiếm, thì bộ máy của SQL Server sẽ phải **duyệt từng dòng** trong bảng.
- Khi dữ liệu của bảng lên tới hàng ngàn, **hàng triệu dòng** và có thể hơn thế nữa, công việc quét(scan) toàn bộ bảng sẽ trở lên **chậm hơn** và làm **tăng thêm các chi phí** (thời gian, tài nguyên...).

# Giới thiệu

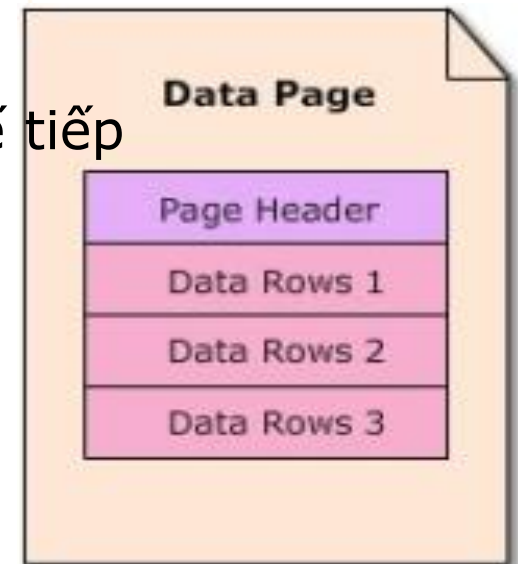
- Vì vậy, mà index được **khuyến khích** nên dùng.
- Việc tạo hay gỡ bỏ các index **không làm ảnh hưởng** **gì đến code** của ứng dụng.
- Các thao tác index là ở phía **backend** (chạy phía sau) với hỗ trợ của bộ máy csdl(database engine).
- Hơn nữa, việc tạo ra index phù hợp có thể làm **tăng** **hiệu suất** đáng kể cho ứng dụng.

# Giới thiệu

- Một quyển sách có nhiều trang, trong mỗi trang có chứa các đoạn văn được tạo nên bởi các câu. Tương tự, SQL Server 2014 lưu trữ dữ liệu trong đơn vị lưu trữ là **data page** (trang dữ liệu).
- SQL Server dùng data page để lưu trữ các dòng dữ liệu của bảng. Mỗi data page có kích thước **8 Kilo Bytes (KB)**.

# Giới thiệu

- Mỗi Mega Byte của csdl sẽ có **128 trang** dữ liệu.
- Mỗi trang được bắt đầu bằng **96 byte header** lưu các thông tin về trang như:
  - **Số trang** (Page number)
  - **Kiểu trang** (Page type)
  - **Không gian** còn trống trên trang
  - **Con trỏ** trỏ đến các trang trước hoặc kế tiếp



# Giới thiệu

- SQL Server sử dụng một trong hai phương pháp sau để tổ chức các trang của bảng:
  1. Các bảng có một clustered index, còn gọi là **bảng clustered**.
  2. Các bảng không có clustered index, còn gọi là **heap**.
- Khi tạo clustered hoặc nonclustered index cho bảng, các **trang index** cũng **được tạo** ra cho bảng để **lưu các khóa**.

# Giới thiệu

- Các trang index giúp cho có thể truy xuất trực tiếp tới dòng bất kỳ trong bảng

## Data Pages

Page 4

Con	...
Funk	...
White	...
...	...
...	...

Page 5

Rudd	...
White	...
Barr	...
...	...
...	...

Page 6

Akhtar	...
Funk	...
Smith	...
Martin	...
...	...

Page 7

Smith	...
Ota	...
Jones	...
...	...
...	...

Page 8

Martin	...
Phua	...
Jones	...
Smith	...
...	...

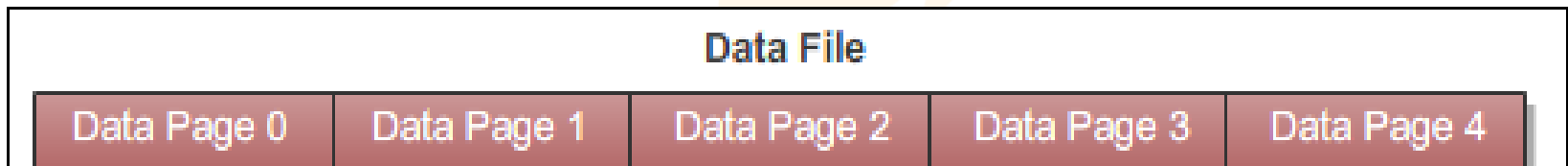
Page 9

Ganio	...
Jones	...
Hall	...
...	...
...	...



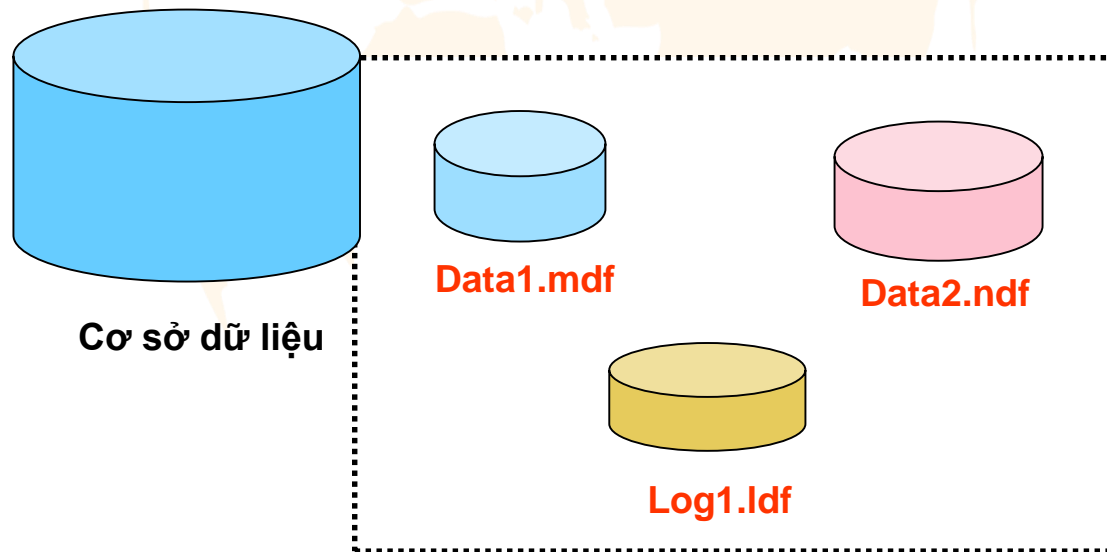
# Giới thiệu

- Tất cả **thao tác nhập/xuất** trong csdl được thực hiện ở **mức trang**. Có nghĩa là bộ máy csdl đọc hoặc ghi các trang dữ liệu.
- Một tập 8 trang dữ liệu(data page) liên tiếp được xem như là một extend. SQL Server 2014 lưu các trang dữ liệu trong các tập tin **data file**.
- Không gian cấp phát cho một data file được chia thành **một dãy các trang dữ liệu** được đánh số bắt đầu từ 0



# Giới thiệu

- Một csdl SQL Server 2014 có ba kiểu tập tin :
  1. Primary Data Files (**.mdf**)
  2. Secondary Data Files (**.ndf**)
  3. Transaction Log Files (**.ldf**)



# SQL Server truy xuất dữ liệu như thế nào?

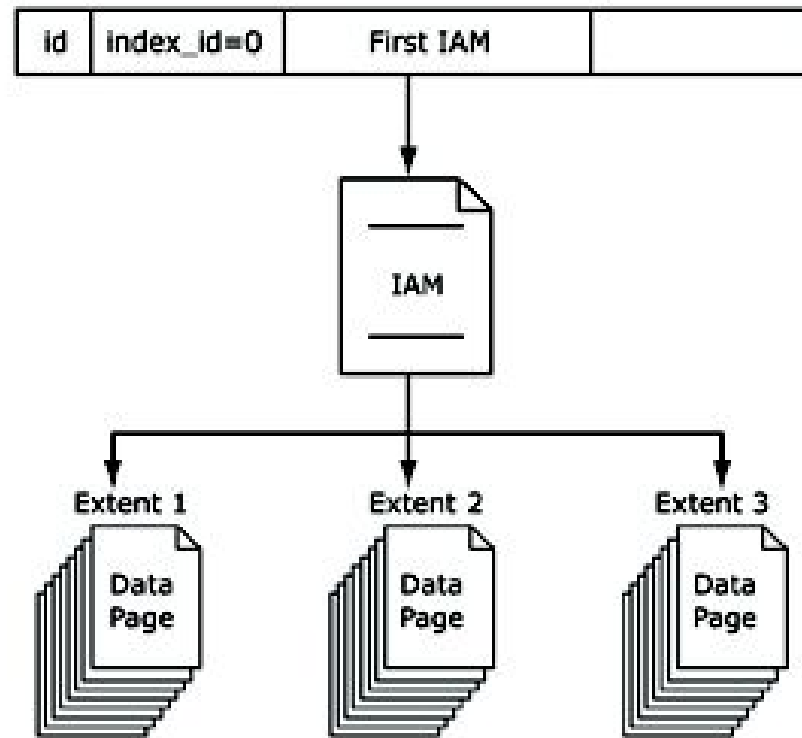
# Giới thiệu

Với các bảng **không có clustered index**. SQL Server truy xuất dữ liệu bằng cách **quét toàn bộ bảng** (table scan). Một bảng được quét như sau:

- Bắt đầu từ **trang dữ liệu đầu tiên** của bảng.
- Quét lần lượt tất cả các dòng **từ trang này tới trang khác** của bảng. **Index Allocation Maps (IAMs)** được sử dụng để di chuyển từ trang này tới trang khác trong một heap (với bảng **có clustered index**, các **liên kết trang** được sử dụng để di chuyển từ trang này tới trang khác).

# Giới thiệu

- **Đọc từng dòng** trong bảng và rút (**extract**) ra các dòng **thỏa mãn điều kiện** của truy vấn.
- Là cách tốt nhất cho việc **truy xuất các bảng nhỏ**.



# Giới thiệu

Với bảng **có clustered index**, SQL Server sử dụng truy xuất index (indexed access) để truy xuất dữ liệu. Truy xuất index được thực hiện như sau:

- **Duyệt cấu trúc cây index** để tìm các dòng mà truy vấn yêu cầu.
- Đọc và lấy(extract) ra các dòng thỏa mãn điều kiện của truy vấn.
- Là cách **tốt nhất để truy xuất** các dòng và dãy các dòng từ một **bảng lớn**.

# Sự cần thiết của Index



# Giới thiệu

- Để thuận tiện, nhanh chóng cho việc lấy dữ liệu từ csdl, SQL Server 2014 cung cấp **tính năng tạo index**.
- Một index trong csdl SQL Server 2014 có chứa các thông tin cho phép bạn tìm dữ liệu được chỉ ra mà không phải quét toàn bộ bảng như hình sau:
- SQL Server sử dụng index **tương tự** như cách chúng ta sử dụng **mục lục của quyển sách**.

Index			
A			
Adapter	1	Border	19
Aggregate	10	Bullet	58
Analysis	13	C	
Average	23	Consistency	20
B		Connect	22
Board	17	Communication	24
Brilliant	18	Character	30

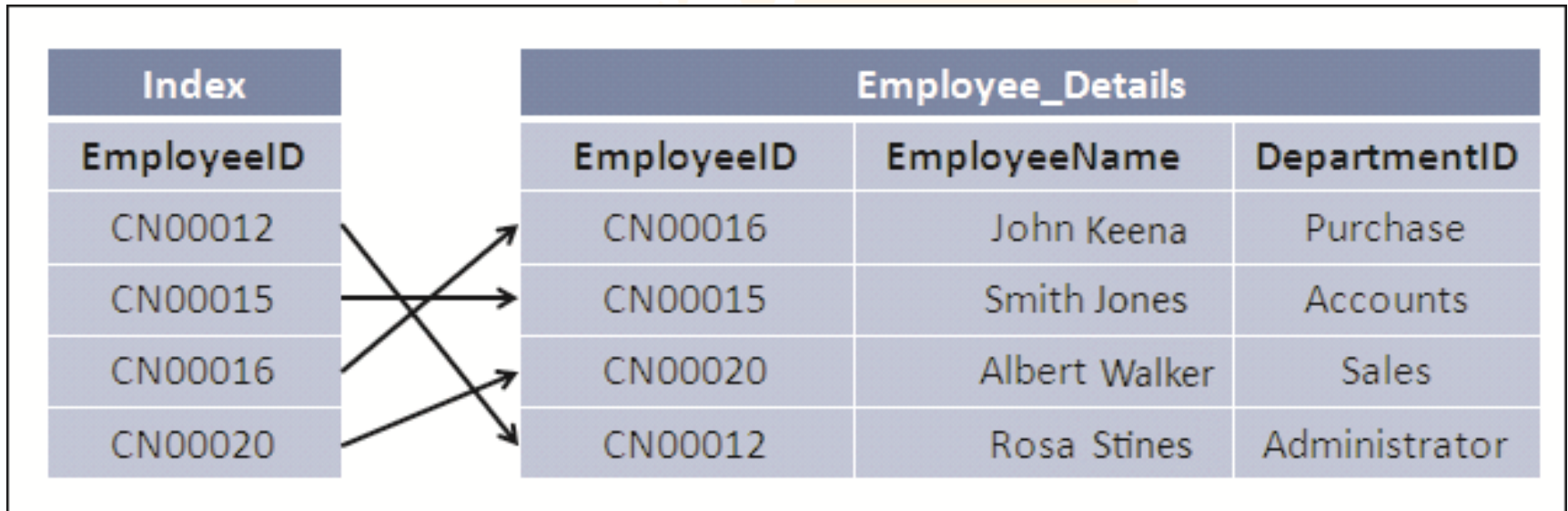


# Giới thiệu

- Trong một bảng, các bản ghi được **lưu trữ theo thứ tự** mà chúng **được nhập vào**. Việc lưu trữ chúng trong cơ sở dữ liệu không được sắp xếp.
- Khi dữ liệu được lấy từ các bảng như vậy, buộc phải **quét (scan) toàn bộ** bảng.
- Điều đó làm cho việc xử lý **truy vấn** lấy dữ liệu **chậm xuống**. Để **tăng tốc độ** truy vấn dữ liệu, các **index** cần phải được tạo.

# Giới thiệu

- Khi index được tạo trên bảng, index tạo ra một thứ tự các dòng dữ liệu trong bảng như hình sau:



# Giới thiệu

- Index được **tạo tự động** khi ràng buộc **PRIMARY KEY** và **UNIQUE** được định nghĩa trên bảng.
- Các index làm **giảm các thao tác** nhập/xuất đĩa và chiếm ít tài nguyên hơn.
- Câu lệnh **CREATE INDEX** được sử dụng để tạo index

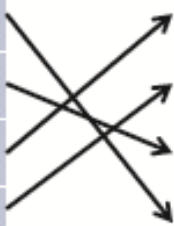
```
CREATE INDEX <tên_index> ON <tên_bảng>(tên_cột)
```

-- Tạo Indexes cho bảng Color trên cột ColorId

```
CREATE INDEX ID_ColorId  
ON Color (ColorId)
```

# Giới thiệu

Hình dưới đây minh họa bảng Customer\_Details được tạo index

Customer_Details				Index	
CustID	AccNo	AccName	Country	IX_Country	
01	CN001	John Keena	Spain		Germany
02	CN020	Smith Jones	Russia		London
03	CN011	Albert Walker	Germany		Russia
04	CN021	Rosa Stines	London		Spain

Các chỉ mục (index) trở tới các vị trí của dòng trên trang dữ liệu thay cho việc tìm kiếm thông qua bảng.

# Hiệu suất

Giải thích kịch bản tìm kiếm nhờ index

- Trong danh mục điện thoại (telephone directory), ở đó một lượng lớn dữ liệu được lưu trữ theo **thứ tự bảng chữ cái** và được **truy xuất thường xuyên**.
- Nếu như dữ liệu đó **không được sắp xếp**, gần như **không thể tìm kiếm** một số điện thoại cụ thể.



# Hiệu suất

- Tương tự, trong csdl có một số lượng lớn các bản ghi thường xuyên được truy cập, dữ liệu nên **được sắp xếp** để việc lấy ra được **nhANH chóng**.
- Khi một index được tạo trên bảng, index sẽ **sắp xếp vật lý hoặc logic** các bản ghi.
- Do vậy, việc **tìm kiếm** các bản ghi trở lên **nhANH hơn**, giảm sử dụng quá mức tài nguyên hệ thống.

# Hiệu suất

Tạo index

- Tăng tốc độ truy xuất dữ liệu
- Đảm bảo tính duy nhất các dòng

## **Nên tạo index:**

Cho khóa chính, khóa ngoại

Thường xuyên tìm kiếm theo một miền

Thường xuyên truy cập theo thứ tự

# Hiệu suất

Không tạo index

- Tiêu tốn không gian đĩa
- Chi phí cao

## **Nên tạo index:**

Cho khóa chính, khóa ngoại

Thường xuyên tìm kiếm theo một miền

Thường xuyên truy cập theo thứ tự



# Kiến trúc

Các bảng của csdl SQL Server 2014 sử dụng một trong hai phương pháp để tổ chức dữ liệu trong các trang dữ liệu trong một phân vùng (partition).

- Nếu bảng có một clustered index, nó được gọi là **bảng clustered**. Các dòng dữ liệu trong bảng được lưu trữ sắp xếp theo thứ tự khóa của clustered index.
- Nếu bảng không có clustered index, được gọi là **bảng heap**. Các dòng dữ liệu trong bảng không được sắp xếp theo thứ tự cụ thể nào.

Employee_Details		
EmpID	EmpName	DeptID
CN00012	Robert	BN0011
CN00018	John	BN0020
CN00019	Smith	BN0021
CN00020	Rosa	BN0001

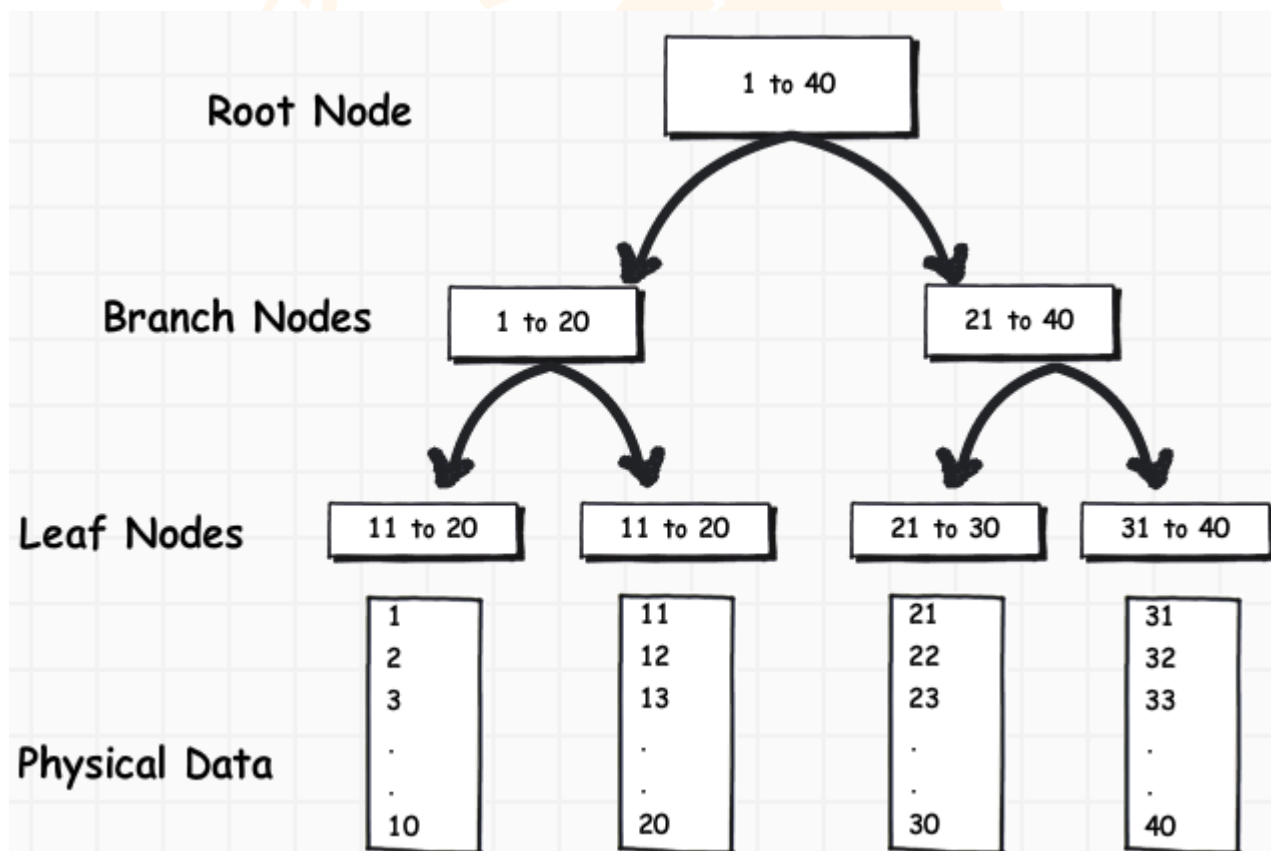
Clustered Structure

Employee_Details		
EmpID	EmpName	DeptID
CN00020	Rosa	BN0001
CN00018	John	BN0020
CN00019	Smith	BN0021
CN00012	Robert	BN0011

Heap Structure

# Kiến trúc

Trong SQL Server, các index được tổ chức theo dạng cây nhị phân (B-tree) như hình dưới:



# Tạo Index

## Clustered Index

- Clustered index được tự động tạo ra khi định nghĩa **khóa chính** cho bảng.
- Clustered index làm cho các bản ghi được sắp xếp ở **mức vật lý** (sắp xếp trong các trang) theo thứ tự của giá trị khóa.

Cú pháp:

```
CREATE CLUSTERED INDEX <tên_index> ON  
<tên_bảng>(tên_cột)
```

# Tạo Index

Ví dụ: tạo clustered index cho bảng Customer

-- Tạo Clustered Indexes cho bảng Customer trên cột  
CustomerId

```
CREATE CLUSTERED INDEX Id_CustomerId ON Customer  
(CustomerId)
```

	CustomerId	CustomerName	Phone	Address	Email	UserName	Password	Created	Status
1	1	Trinh Dinh Long	0956783453	Hai Duong	long@gmail.com	Dinh Long	long123	2017-08-26	1
2	2	Hoang Tuan Anh	087543178	Nam Dinh	anh@gmail.com	Tuan Anh	123	2017-08-26	1
3	3	Do My Linh	0989654387	Hai Phong	linhdm@gmail.com	Linh Do	123456	2017-05-26	1
4	4	Tran The tai	0865432098	Ha Nam	tai@gmail.com	Tran Tai	abcd	2017-05-09	1
5	5	Nguyen Tien Thanh	01654325677	Thai Nguyen	tienthanh@gmail.com	Tien Thanh	ththth	2017-07-09	1
6	6	Trinh Dinh Quang	0966678456	Thanh Hoa	quang@gmail.com	Quang Di...	quang567	2017-08-08	1

# Tạo Index

Lưu ý:

- Cột được chọn để tạo clustered index cho bảng phải là cột **không có giá trị trùng nhau**.
- Index này sẽ **tổ chức** các bản ghi **theo thứ tự** của các **giá trị trong cột index**.
- Clustered index được **dùng để định vị** (tìm kiếm) một dòng hoặc một dãy các dòng.
- Bắt đầu từ trang đầu tiên của cây index, giá trị tìm được **so sánh với mỗi khóa** (key) trên trang.

# Tạo Index

Khi tạo clustered index, hãy xem xét các hướng dẫn sau đây:

- Chỉ có thể tạo **một clustered index** cho mỗi bảng.
- **Thứ tự vật lý** các dòng trong bảng và **thứ tự các dòng** trong index là **như nhau**.
- Các giá trị khóa trong một clustered index phải là **duy nhất**.

# Tạo Index

Trong trường hợp bảng **không có khóa chính**, Clustered Index nên được xem xét các cột sau:

- Các cột khóa **được tìm kiếm** rộng rãi.
- Các cột được sử dụng trong các truy vấn mà trả lại **tập kết quả lớn**.
- Các cột có **dữ liệu duy nhất**.
- Các cột được sử dụng việc **join các bảng**.

# Tạo Index

## Nonclustered Index

- Nonclustered index được định nghĩa trên bảng có dữ liệu được tổ chức theo **cấu trúc clustered** hoặc **cấu trúc heap**.
- Nonclustered index là loại **index mặc định** nếu không có index được định nghĩa trên bảng.
- Mỗi dòng trong nonclustered index có chứa một giá trị khóa nonclustered và bộ định vị dòng(row locator)
- **Bộ định vị dòng** trỏ tới dòng dữ liệu trong bảng tương ứng với giá trị khóa.



# Tạo Index

Cú pháp:

```
CREATE NONCLUSTERED INDEX <tên_index> ON  
<tên_bảng> (tên_cột)
```

-- Tạo NonClustered Indexes cho bảng Customer trên cột  
Email

```
CREATE NONCLUSTERED INDEX NID_Email_Customer ON  
Customer (Email)
```

	CustomerId	CustomerName	Phone	Address	Email	UserName	Password	Created	Status
1	1	Trinh Dinh Long	0956783453	Hai Duong	long@gmail.com	Dinh Long	long123	2017-08-26	1
2	2	Hoang Tuan Anh	087543178	Nam Dinh	anh@gmail.com	Tuan Anh	123	2017-08-26	1
3	3	Do My Linh	0989654387	Hai Phong	linhdm@gmail.com	Linh Do	123456	2017-05-26	1
4	4	Tran The tai	0865432098	Ha Nam	tai@gmail.com	Tran Tai	abcd	2017-05-09	1
5	5	Nguyen Tien Thanh	01654325677	Thai Nguyen	tienthanh@gmail.com	Tien Thanh	ththth	2017-07-09	1
6	6	Trinh Dinh Quang	0966678456	Thanh Hoa	quang@gmail.com	Quang Di...	quang567	2017-08-08	1

# Tạo Index

Clustered Index	Nonclustered Index
Sử dụng cho các truy vấn trả về các tập kết quả lớn	Sử dụng cho các truy vấn không trả về các tập kết quả lớn
Chỉ có thể tạo một clustered index cho một bảng	Có thể tạo nhiều nonclustered index cho bảng
Dữ liệu được sắp xếp theo thứ tự trên khóa clustered	Dữ liệu không được sắp xếp theo thứ tự trên khóa nonclustered
Các nút lá của một clustered index có chứa các trang dữ liệu (data page).	Các nút lá của một nonclustered index có chứa các trang chỉ mục(index page).

# Tóm tắt bài học

- Index **hỗ trợ tìm kiếm dữ liệu** trên bảng (có số lượng bản ghi rất lớn) với thời gian nhanh chóng.
- Một bảng có tối đa **1 clustered index** và **999 nonclustered index**.
- Trong khi xóa tất cả index trên bảng, trước tiên bạn phải **xóa nonclustered index**, sau đó là **clustered indexes**.

## TRƯỜNG ĐÀO TẠO LẬP TRÌNH VIÊN VÀ QUẢN TRỊ MẠNG QUỐC TẾ BACHKHOA-APTECH

---

# Thank for watching!

