# Database

# Partition

# 1. Khi nào thì đánh partition trong 1 bảng?

### **1. Khi có tập dữ liệu lớn**

* **Kích thước bảng lớn**: Nếu bạn có một bảng với số lượng bản ghi rất lớn (ví dụ, hàng triệu hoặc hàng tỷ bản ghi), partitioning có thể giúp cải thiện hiệu suất truy vấn. Việc chia nhỏ bảng thành các phân vùng giúp DBMS xử lý và quét dữ liệu nhanh hơn.

### **2. Khi cần cải thiện hiệu suất truy vấn**

* **Tăng tốc độ truy vấn**: Khi các truy vấn thường xuyên chỉ truy cập một phần nhỏ của dữ liệu (ví dụ: các truy vấn theo thời gian), việc phân vùng có thể giúp tăng tốc độ thực thi bằng cách giới hạn lượng dữ liệu mà hệ thống phải quét. Ví dụ, nếu bạn thường xuyên truy vấn dữ liệu theo tháng hoặc năm, phân vùng theo ngày hoặc tháng sẽ có lợi.

### **3. Khi thực hiện bảo trì dữ liệu thường xuyên**

* **Dễ dàng bảo trì và quản lý**: Khi bạn cần thực hiện các thao tác bảo trì như sao lưu, phục hồi, hoặc xóa dữ liệu cũ, việc phân vùng sẽ giúp đơn giản hóa quy trình. Bạn có thể dễ dàng xóa hoặc lưu trữ toàn bộ phân vùng mà không làm ảnh hưởng đến toàn bộ bảng.

### **4. Khi cần cải thiện hiệu suất ghi**

* **Phân bổ tải ghi**: Với các ứng dụng có lưu lượng ghi cao, partitioning có thể giúp giảm tải cho các hoạt động ghi bằng cách phân phối dữ liệu qua nhiều phân vùng. Điều này giúp giảm thiểu khả năng khóa (locking) và cải thiện hiệu suất ghi.

### **5. Khi có yêu cầu bảo mật và quản lý quyền truy cập**

* **Quản lý quyền truy cập khác nhau**: Nếu bạn có yêu cầu về bảo mật hoặc phân quyền truy cập khác nhau cho các nhóm người dùng, phân vùng có thể giúp bạn kiểm soát quyền truy cập tốt hơn bằng cách áp dụng quyền truy cập cho từng phân vùng.

### **6. Khi có dữ liệu không đồng nhất**

* **Dữ liệu không đồng nhất**: Khi dữ liệu có các đặc điểm khác nhau (ví dụ: các loại dữ liệu, nguồn dữ liệu), phân vùng có thể giúp tổ chức và tối ưu hóa cách mà dữ liệu được lưu trữ và truy vấn.

### **7. Khi dữ liệu có xu hướng tăng theo thời gian**

* **Dữ liệu theo thời gian**: Nếu dữ liệu của bạn liên tục tăng theo thời gian (ví dụ: nhật ký giao dịch), bạn có thể phân vùng theo thời gian (ngày, tháng, năm) để dễ dàng quản lý và truy vấn.

### **8. Khi cần quản lý dữ liệu theo ngày hoặc theo tuần**

* **Dữ liệu theo thời gian**: Nếu dữ liệu được ghi nhận theo ngày hoặc theo tuần và bạn cần truy xuất hoặc phân tích dữ liệu theo các khoảng thời gian cụ thể, việc phân vùng sẽ giúp dễ dàng truy xuất và quản lý dữ liệu theo từng khoảng thời gian.

# 2. Đánh partition mang lại lợi ích gì cho truy vấn?

### **1. Tăng tốc độ truy vấn**

* **Giảm lượng dữ liệu quét**: Khi một bảng được phân vùng, DBMS có thể chỉ quét các phân vùng cần thiết thay vì quét toàn bộ bảng. Điều này giảm thiểu thời gian truy vấn, đặc biệt là trong các trường hợp truy vấn dữ liệu theo điều kiện phân vùng, chẳng hạn như theo thời gian (ngày, tháng).

### **2. Cải thiện hiệu suất truy vấn theo thời gian**

* **Truy vấn theo khoảng thời gian**: Nếu dữ liệu được phân vùng theo thời gian (ví dụ: theo tháng hoặc năm), các truy vấn chỉ cần truy cập một phân vùng cụ thể, giúp tăng tốc độ truy vấn và giảm thời gian chờ đợi.

### **3. Tối ưu hóa việc sử dụng bộ nhớ**

* **Tối ưu hóa bộ nhớ đệm**: Phân vùng có thể cải thiện hiệu suất của bộ nhớ đệm (cache). DBMS có thể giữ các phân vùng được truy cập thường xuyên trong bộ nhớ đệm, làm tăng tốc độ truy vấn mà không cần phải đọc dữ liệu từ đĩa.

### **4. Dễ dàng bảo trì và quản lý dữ liệu**

* **Bảo trì linh hoạt**: Việc bảo trì dữ liệu, chẳng hạn như xóa hoặc lưu trữ các phân vùng cũ, trở nên dễ dàng hơn. Bạn có thể xóa một phân vùng hoàn toàn mà không làm ảnh hưởng đến các phân vùng khác, giúp cải thiện hiệu suất tổng thể.

### **5. Giảm thiểu khóa (Locking)**

* **Giảm thiểu khóa**: Khi một truy vấn yêu cầu chỉ đọc một phân vùng cụ thể, việc này có thể giảm thiểu khóa (locking) trên toàn bộ bảng, giúp tăng cường khả năng đồng thời của hệ thống (concurrency).

### **6. Quản lý dữ liệu hiệu quả**

* **Tổ chức và quản lý dữ liệu**: Phân vùng giúp tổ chức dữ liệu theo cách dễ dàng hơn cho việc truy vấn và phân tích. Bạn có thể chia dữ liệu theo các tiêu chí cụ thể (như khu vực, loại dữ liệu, v.v.), giúp truy vấn trở nên rõ ràng và có tổ chức hơn.

### **7. Cải thiện hiệu suất ghi**

* **Tăng tốc độ ghi**: Đối với các ứng dụng có lượng ghi lớn, phân vùng có thể giúp phân phối các hoạt động ghi qua nhiều phân vùng, giảm thiểu khả năng khóa và cải thiện hiệu suất ghi.

### **8. Tối ưu hóa kế hoạch thực thi**

* **Tối ưu hóa kế hoạch thực thi**: DBMS có thể tạo ra kế hoạch thực thi tối ưu hơn khi biết rằng dữ liệu được phân chia thành các phân vùng. Điều này giúp giảm thời gian thực thi cho các truy vấn phức tạp.