


ChatGPT - SQL Expert (QueryGPT)

 chatgpt.com/g/g-m5lMeGifF-sql-expert-querygpt/c/682c7f84-c6dc-8006-8567-84f3dddf4464



Используется Bitmap Index Scan используется при какой селективности и кардинальности?

Bitmap Index Scan используется в PostgreSQL и других СУБД при **средней селективности** запроса — то есть когда:

- **Селективность**: не слишком мала (как у уникального индекса), но и не слишком велика (как при полном сканировании таблицы). Обычно — от **1% до 10–20%** от общего количества строк таблицы.
- **Кардинальность**: может быть как высокая, так и средняя. Особенно полезен при **комбинации нескольких условий** по разным индексам — в этом случае PostgreSQL может построить несколько bitmap-сканов и объединить их (**BitmapAnd**, **BitmapOr**).

Подробности

- Если **селективность очень высокая** (например, условие возвращает >30% строк), PostgreSQL часто предпочитает **Seq Scan** (полное сканирование).
- Если **селективность очень низкая** (например, <0.1%), чаще выбирается **Index Scan**, особенно если нужно вернуть несколько строк.
- **Bitmap Index Scan** эффективен, когда нужно извлечь **много строк, но не все**, и при этом можно **выгодно использовать индекс**.

Пример:

sql

```
SELECT * FROM orders WHERE status = 'shipped';
```

Если `status` принимает всего 3–4 значения, и `'shipped'` — это ~10% строк, то `Bitmap Index Scan` будет уместным выбором.

Follow GPT creator profile on LinkedIn <https://linkedin.com/in/khanukov/>

Как здесь увязана кардинальность?

Отличный вопрос! Давайте разберём, как кардинальность влияет на выбор `Bitmap Index Scan`.

♦ Кардинальность: что это?

Кардинальность — это количество уникальных значений в колонке. Примеры:

- Низкая кардинальность: колонка `status` со значениями `['new', 'shipped', 'cancelled']`.
- Высокая кардинальность: колонка `user_id`, где почти каждое значение уникально.

♦ Связь с выбором `Bitmap Index Scan`

- Низкая кардинальность: `Bitmap Index Scan` полезен, когда запрос выбирает все строки с определённым значением, встречающимся во многих строках. Пример:

sql

```
SELECT * FROM events WHERE event_type = 'login';
```

`event_type` имеет 5 уникальных значений — отличная ситуация для `bitmap`-скана, если `'login'` встречается в, скажем, 15% строк.

- **Высокая кардинальность:** в этом случае чаще используется **обычный Index Scan**, потому что значения более селективны. Например:

sql

```
SELECT * FROM users WHERE user_id = 123;
```

Здесь предпочтительнее **Index Scan**, так как возвращается только одна строка.



Ключевая идея

- **Bitmap Index Scan** полезен при **НИЗКОЙ** кардинальности и **СРЕДНЕЙ** селективности.
- При **высокой кардинальности** и малом числе совпадений лучше работает **Index Scan**.
- При **низкой селективности** (>30%) даже при низкой кардинальности будет предпочтён **Seq Scan**.

Follow GPT creator profile on LinkedIn <https://linkedin.com/in/khanukov/>

2/2