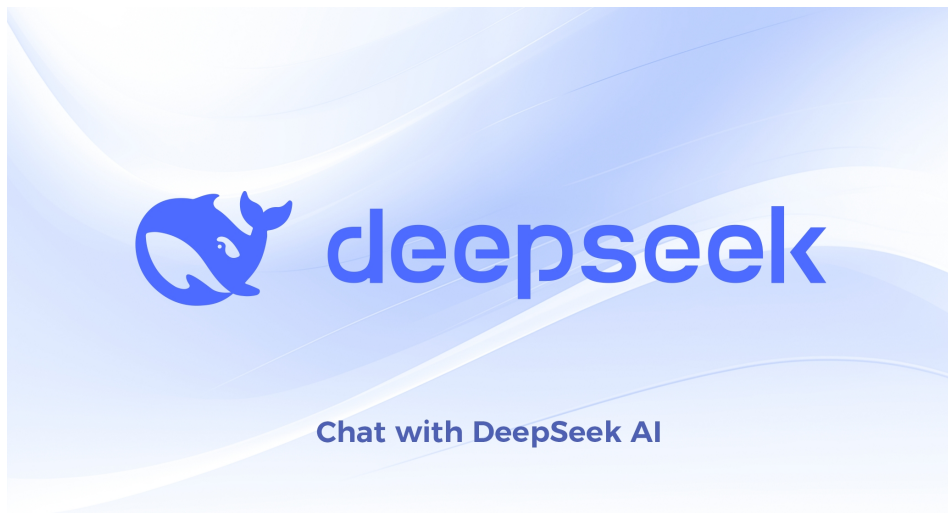


# DeepSeek - Into the Unknown

---

 [chat.deepseek.com/a/chat/s/dc697c73-6dfe-4b79-97d3-1c2caa48efc3](https://chat.deepseek.com/a/chat/s/dc697c73-6dfe-4b79-97d3-1c2caa48efc3)



Команда **DISTINCT** в PostgreSQL используется для устранения дубликатов строк из результата запроса. Вот как она работает:

## Основной алгоритм работы

---

### 1. Сортировка и группировка (основной метод):

- PostgreSQL сначала выполняет основной запрос
- Затем результаты сортируются (аналогично **ORDER BY** по всем столбцам)
- После сортировки удаляются соседние дубликаты

### 2. Хеширование (альтернативный метод для больших наборов данных):

- Создается хеш-таблица, где ключ - это комбинация значений всех столбцов
- Для каждой строки вычисляется хеш и проверяется наличие в таблице
- Если хеш уже существует, строка пропускается

## Оптимизации

---

PostgreSQL может выбрать между этими методами в зависимости от:

- Размера результата
- Доступной памяти
- Наличия индексов

Если в запросе есть **ORDER BY**, PostgreSQL может использовать уже отсортированные данные для **DISTINCT**.

## Примеры

---

sql

Copy

Download

```
-- Базовое использование
SELECT DISTINCT department FROM employees;

-- DISTINCT по нескольким столбцам
SELECT DISTINCT department, job_title FROM employees;
```

## Особенности

---

1. **DISTINCT** применяется ко всем столбцам в SELECT
2. Для частичного DISTINCT используйте **DISTINCT ON**:

sql

Copy

Download

```
-- Одна строка для каждого отдела с самым высоким окладом
SELECT DISTINCT ON (department) department, employee_name, salary
FROM employees
ORDER BY department, salary DESC;
```

3. **DISTINCT** может быть ресурсоемким для больших таблиц, так как требует полной обработки данных перед удалением дубликатов.

Оптимизатор PostgreSQL автоматически выбирает наиболее эффективный метод выполнения **DISTINCT** для каждого конкретного запроса.