

Substrings

Substrings – Grundlagen, Anwendung und Praxis

Einführung

Ein **Substring** ist ein Teil einer Zeichenkette, der durch Angabe eines Startpunkts und einer Länge oder durch ein Muster extrahiert wird. Substrings sind essenziell, wenn es darum geht, Daten aus komplexen Zeichenketten zu extrahieren oder diese zu manipulieren.

In diesem Thema lernst du:

1. Was Substrings sind und wofür sie verwendet werden.
2. Wie Substrings definiert und extrahiert werden.
3. Praktische Beispiele zur Anwendung von Substrings in PostgreSQL.

Grundlagen von Substrings

Ein Substring ist eine Teilzeichenkette, die durch:

1. Startposition und Länge
2. oder durch ein Muster definiert wird.

Beispiel:

- Zeichenkette: `Hello World`
- Substring (Start bei 1, Länge 5): `Hello`

Substrings in PostgreSQL

Die Funktion `substring()` in PostgreSQL bietet eine einfache Möglichkeit, Substrings zu extrahieren. Die allgemeine Syntax lautet:

1. Mit Position und Länge

```
1 substring(string FROM start_position FOR length)
2
```

- `string` : Die Zeichenkette, die verarbeitet wird.
- `start_position` : Die Startposition (1-basiert).



- `length` : Die Anzahl der zu extrahierenden Zeichen.

2. Mit regulären Ausdrücken

SQL

```
1 substring(string FROM 'regex_pattern')
2
```

- `regex_pattern` : Ein regulärer Ausdruck, der definiert, welcher Teil der Zeichenkette extrahiert wird.

Praxisbeispiele

Beispiel 1: Extraktion basierend auf Startposition und Länge

Angenommen, du hast eine Zeichenkette `ABCDEFGH`, und möchtest die Zeichen von der dritten bis zur fünften Position extrahieren:

SQL

```
1 SELECT substring('ABCDEFGH' FROM 3 FOR 3) AS result;
2 -- Ergebnis: 'CDE'
3
```

Beispiel 2: Dynamische Substring-Extraktion mit `position()`

Wenn die Position eines Zielzeichens innerhalb der Zeichenkette variieren kann, kannst du die Funktion `position()` verwenden, um die Startposition dynamisch zu bestimmen:

Extrahiere den Teil vor dem Bindestrich (-) in der Zeichenkette `2024-Q1-Report` :

SQL

```
1 SELECT substring('2024-Q1-Report' FROM 1 FOR
   position('-' IN '2024-Q1-Report') - 1) AS result;
2 -- Ergebnis: '2024'
3
```

Beispiel 3: Substrings mit regulären Ausdrücken

Extrahiere die Domain aus einer E-Mail-Adresse:

SQL

```
1 SELECT substring('user@example.com' FROM '@(.*?)$') AS
   domain;
```

```
2  -- Ergebnis: 'example.com'
3
```

Beispiel 4: Letzte Zeichen extrahieren

Um die letzten `n` Zeichen einer Zeichenkette zu extrahieren, verwende `length()` :

```
SQL
1  SELECT substring('ABCDEFGH' FROM length('ABCDEFGH') -
2    2 + 1 FOR 3) AS result;
2  -- Ergebnis: 'FGH'
3
```

Häufige Anwendungsfälle

1. **Datenbereinigung:** Entferne irrelevante Präfixe oder Suffixe aus Strings.

```
SQL
1  SELECT substring('ID:12345' FROM 4) AS id_value;
2  -- Ergebnis: '12345'
3
```

2. **Code-Splitten:** Teile komplexe Codes in ihre Bestandteile auf.

```
SQL
1  SELECT substring('ABC-123-XYZ' FROM 5 FOR 3) AS
   code_part;
2  -- Ergebnis: '123'
3
```

3. **Datenanalyse:** Extrahiere dynamische Informationen basierend auf Mustern.

```
SQL
1  SELECT substring('Order:456' FROM ':(\\d+)$') AS
   order_number;
2  -- Ergebnis: '456'
3
```

Übungen

1. Extrahiere die ersten 4 Zeichen aus der Zeichenkette 'PGSQL-2024'.
2. Schreibe eine Abfrage, um alles nach dem ersten Doppelpunkt (:) in der Zeichenkette 'DATA:CONTENT' zu extrahieren.
3. Extrahiere die letzten 5 Zeichen aus einer beliebigen Zeichenkette.

Zusammenfassung

Substrings sind ein mächtiges Werkzeug zur Zeichenkettenmanipulation. Durch die Kombination von festen Positionen, Längen und regulären Ausdrücken kannst du flexibel und effizient Teile von Zeichenketten extrahieren.

