Complete SQL for Data Science Cheatsheet

Introduction

This cheatsheet provides a comprehensive overview of SQL concepts for data science. It is designed to assist beginners in understanding and implementing SQL queries for data manipulation and analysis.

Data Manipulation

Creating a Database

To create a new database in SQL, use the following code:

```
CREATE DATABASE database_name;
```

Creating a Table

To create a new table in SQL, use the following code:

```
CREATE TABLE table_name (
    column1 datatype1,
    column2 datatype2,
    column3 datatype3,
    ...
);
```

Inserting Data

To insert data into a table in SQL, use the following code:

```
INSERT INTO table_name (column1, column2, column3, ...)
VALUES (value1, value2, value3, ...);
```

Querying Data

To query data from a table in SQL, use the following code:

```
SELECT column1, column2, ...
FROM table_name
WHERE condition;
```

Updating Data

To update data in a table in SQL, use the following code:

```
UPDATE table_name
SET column1 = value1, column2 = value2, ...
WHERE condition;
```

Deleting Data

To delete data from a table in SQL, use the following code:

```
DELETE FROM table_name
WHERE condition;
```

Data Analysis

Aggregating Data

To perform data aggregation in SQL, use the following code:

```
SELECT column1, AGGREGATE_FUNCTION(column2)
FROM table_name
GROUP BY column1;
```

Joining Tables

To join multiple tables in SQL, use the following code:
lstlisting[language=SQL] SELECT column1, column2, ... FROM table1 JOIN table2 ON table1.column = table2.column;

Sorting Data

To sort data in SQL, use the following code:

```
SELECT column1, column2, ...
FROM table_name
ORDER BY column1 ASC, column2 DESC;
```

Filtering Data

To filter data in SQL, use the following code:

```
SELECT column1, column2, ...
FROM table_name
WHERE condition;
```

Grouping Data

To group data in SQL, use the following code:

```
SELECT column1, AGGREGATE_FUNCTION(column2)
FROM table_name
GROUP BY column1;
```

Subqueries

To use subqueries in SQL, use the following code:

```
SELECT column1, column2, ...
FROM table1
WHERE column1 IN (SELECT column1 FROM table2 WHERE condition);
```

Views

To create a view in SQL, use the following code:

```
CREATE VIEW view_name AS
SELECT column1, column2, ...
FROM table_name
WHERE condition;
```

Advanced SQL

Database Joins

SQL supports different types of joins:

- Inner Join: Retrieves records that have matching values in both tables.
- Left Join: Retrieves all records from the left table and matching records from the right table.
- Right Join: Retrieves all records from the right table and matching records from the left table.
- Full Outer Join: Retrieves all records when there is a match in either the left or right table.

SQL Functions

SQL provides various functions for data manipulation and analysis, including:

- Aggregation Functions: SUM, AVG, MIN, MAX, COUNT.
- String Functions: CONCAT, SUBSTRING, LENGTH, UPPER, LOWER.
- Date Functions: DATE, YEAR, MONTH, DAY, HOUR, MINUTE, SECOND.
- Mathematical Functions: ABS, ROUND, CEIL, FLOOR, POWER, SQRT.

SQL Indexes

Indexes in SQL improve the performance of queries by allowing faster data retrieval. To create an index, use the following code:

```
CREATE INDEX index_name ON table_name (column1, column2, ...);
```

SQL Transactions

Transactions in SQL ensure data integrity and consistency. Use the following code to start a transaction:

```
START TRANSACTION;
```

To commit a transaction:

```
COMMIT;
```

To rollback a transaction:

```
ROLLBACK;
```

Conclusion

This cheatsheet provides a comprehensive overview of SQL concepts for data science. It covers data manipulation, data analysis, advanced SQL techniques, and important functions. With this cheatsheet, beginners can effectively use SQL for data manipulation and analysis tasks in their data science projects.