

Programming Foundation for Data Science

Basic SQL

SQL for Data Science

SQL (Structured Query Language) merupakan sebuah **bahasa** dan digunakan untuk menarik data atau memanipulasi data dari sebuah database dengan tipe **RDBMS**.

Contoh Manipulasi Data di RDBMS

1. Menambah data (*Create*)
2. Membaca data (*Read*)
3. Mengubah data (*Update*)
4. Menghapus data (*Delete*)

Karakteristik Tabel di SQL

1. Nama tabel biasanya ditulis dengan huruf kecil
2. Spasi pada nama tabel biasanya diganti dengan *underscore* (_)
3. Setiap tabel memiliki *field* atau *header*
4. Satu baris horizontal pada tabel disebut *row* atau *record*
5. Satu kolom vertikal pada tabel disebut *column*

Macam-macam Fungsi dan Keyword SQL

1. SELECT, FROM, WHERE, LIMIT
2. ORDER BY
3. Min dan Max
4. Sum dan Average
5. GROUP BY

Macam-macam Syntax SQL

1. Syntax **Union**: Menggabungkan isi dari beberapa tabel

2. Syntax **Join**: Menggabungkan kolom-kolom pada tabel

Script Documentation and Collaboration

Github

Github memberikan solusi **kolaboratif** dan **versioning** bagi para **data scientist** yang bekerja dalam sebuah tim yang besar.

Cara Registrasi Github

1. Kunjungi www.github.com dan klik *sign up*
2. Masukkan *username*, *e-mail address*, dan *password*
3. Verifikasi akun melalui *e-mail*
4. *Sign in* Github

Cara Membuat Repository Baru di Github

1. Klik *icon plus (+)* di kanan atas
2. Pilih *new repository*
3. Membuat *repository name*

Cara Menyalin Repository ke Device

1. Kunjungi www.git-scm.com
2. Pilih Mac OS X, Windows atau Linux/Unix sesuai *device*
3. Kembali ke halaman *repository* di Github
4. Salin *link repository* di Github
5. Membuat folder baru untuk *project*
6. Setelah membuat folder, jalankan *command* (**git init**)
7. Salin *project* dari Git ke dalam *device* (**git clone <link repository>**)

R Language

R Language

R merupakan sebuah **bahasa pemrograman** dan **environment** yang biasa digunakan untuk **statistical computing** atau **graphics** dan dirilis pada tahun 1997.

| No | Kelebihan R Language |
|----|--|
| 1 | R adalah <i>software</i> yang gratis atau <i>open source</i> . |
| 2 | R dapat digunakan diberbagai <i>operating system</i> , seperti <i>Windows</i> , <i>Linux</i> , dan <i>Mac OS</i> sehingga <i>source code</i> dapat dijalankan di <i>operating system</i> dan di <i>platform</i> manapun. |
| 3 | Instalasi dan konfigurasi R yang mudah. |
| 4 | <i>Coding</i> di R cukup mudah dibandingkan <i>statistical tools</i> yang lain. |
| 5 | <i>Packages</i> dan <i>library</i> R sudah banyak dikembangkan. |

| No | Kekurangan R Language |
|----|---|
| 1 | <i>Management memory</i> di R yang tidak terlalu bagus. |
| 2 | Pemrosesannya cukup lambat jika melakukan banyak <i>looping</i> . |

| No | Fitur R dalam Pengolahan Data |
|----|---|
| 1 | Melalui R bisa mengambil data dari berbagai tipe dan format. |
| 2 | R menyediakan berbagai <i>function</i> yang dikemas dalam sebuah <i>package</i> untuk melakukan berbagai eksplorasi data. |
| 3 | Bisa melakukan visualisasi data melalui R dengan berbagai <i>plot/chart/grafik</i> . |
| 4 | Melalui R bisa mengembangkan model <i>machine learning</i> menggunakan data yang sudah dimiliki. |

Python Language

Python Language

Python merupakan sebuah **bahasa pemrograman** yang **general purpose**, yakni dapat digunakan untuk membuat *web*, *software*, *game*, analisis data, maupun komputasi yang rumit.

Aspek Penting Python Bagi *Data Scientist*:

- ✓ Sebagai bahasa pemrograman yang *general purpose*, Python membantu pekerjaan *data scientist* yang kompleks, seperti membuat *query*, *scraping* data dari web, membersihkan data, membuat visualisasi, membangun model, hingga melakukan *deployment* ke *production*.
- ✓ Python memiliki *library* dalam mendukung pekerjaan *data scientist* yang membutuhkan banyak *tools* seperti *pandas*, *numpy*, *scipy*, serta *scikit learn* yang populer dipakai untuk membangun model *machine learning*.

| No | Kelebihan Python Language |
|----|---|
| 1 | Python merupakan bahasa pemrograman yang <i>general purpose</i> . |
| 3 | Syntax Python mudah dibaca karena menggunakan bahasa inggris. |