

Virtual Internship Experience

id/x partners

**Data Science
Tools**

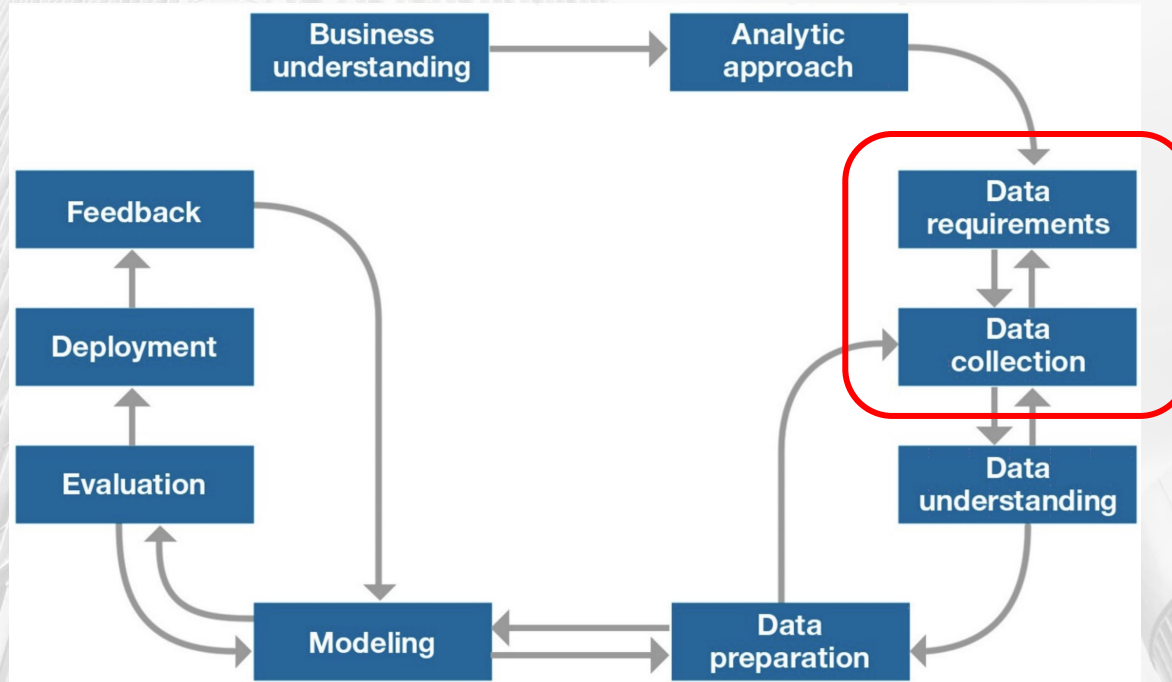
Disclaimer

“Dokumen ini memiliki hak cipta. Barang siapa yang menyebarluaskan atau menduplikasi tanpa izin dari instansi terkait dapat diproses sesuai dengan ketentuan hukum yang berlaku.”

Outline

- 1 **Pengumpulan Data**
- 2 **Bahasa Pemrograman**
- 3 **Version Control**
- 4 **Penggunaan Tools Data Science di ID/X Partners**

Metodologi Data Science





Pengumpulan Data:

Sumber Data / *Data Sources*

Data dapat berasal dari berbagai macam sumber:

- 1. Flat files (.txt, .csv, dll)**
- 2. File Excel / Google Spreadsheet**
- 3. Relational Database (format tabular)**
- 4. Unstructured data (gambar, audio, dll)**
- 5. ...**

Juga menjadi salah satu tools penyimpanan data yang populer.

Pengumpulan Data: RDBMS / SQL

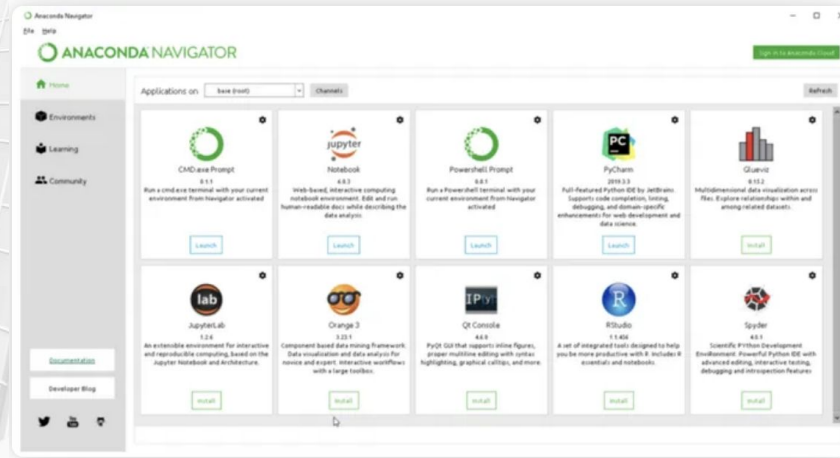
Data yang diambil dari berbagai macam sumber dapat disimpan ke dalam suatu Database Relasional (RDBMS), kemudian dapat dimanipulasi lebih lanjut menggunakan SQL.

```
1  SELECT  m2.nama_kategori, m1.nama_produk,  
2          tl.tgl_transaksi, m3.nama_cabang,  
3          tl.nilai_jual, tl.jumlah_unit  
4  FROM transaksi tl  
5  JOIN ms_produk m1 ON tl.kode_produk = m1.kode_produk  
6  JOIN ms_kategori m2 ON m1.kode_kategori = m2.kode_kategori  
7  JOIN ms_cabang m3 ON m3.kode_cabang = tl.kode_cabang;  
8
```

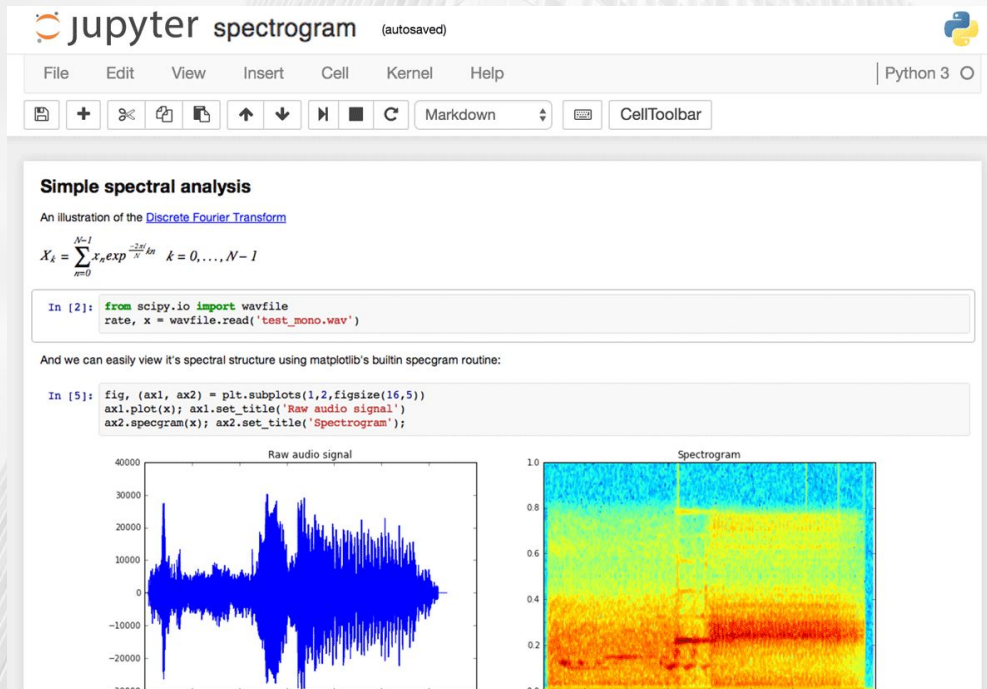


Bahasa Pemrograman: Anaconda

Anaconda merupakan sebuah platform Data Science berupa distribusi bahasa pemrograman Python dan R, bertujuan untuk mempermudah instalasi dan *package management*.



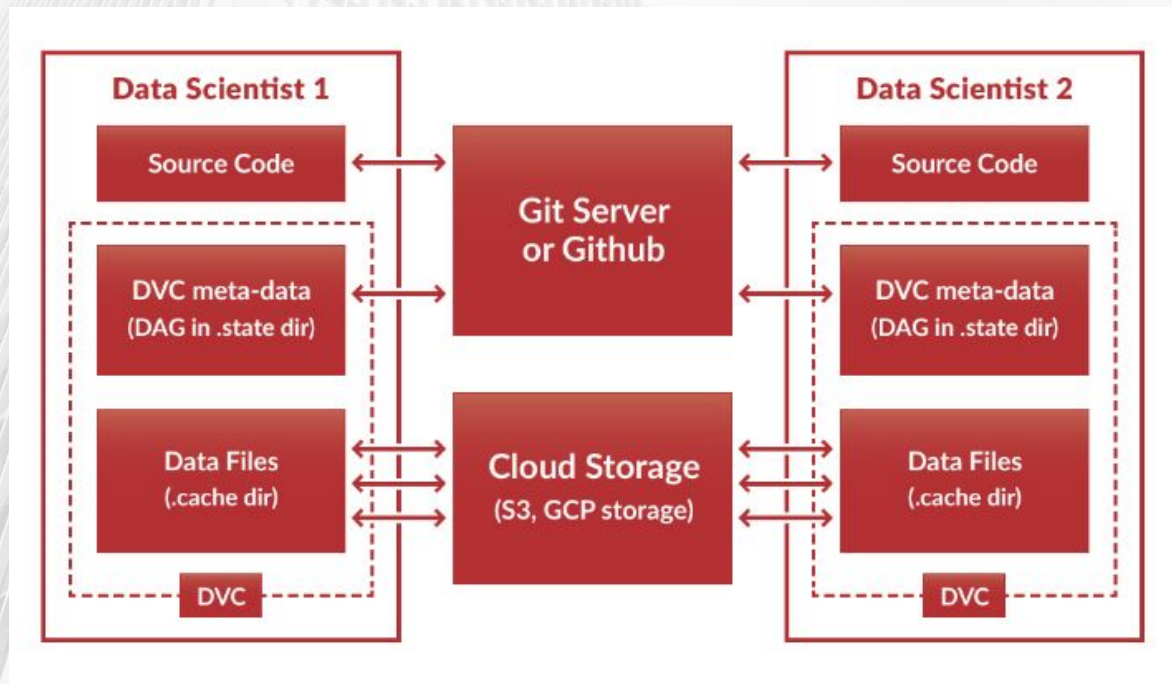
Bahasa Pemrograman: Jupyter Notebook - Python/R



Jupyter Notebook adalah IDE yang populer digunakan oleh Data Scientist untuk melakukan analisis data, visualisasi, membuat data pipeline, bahkan membuat model machine learning sekalipun dengan menggunakan bahasa pemrograman Python ataupun R.

Deployment:

Version Control - Git



Deployment:

Version Control - Git

Git adalah sebuah perangkat untuk mengatur segala macam perubahan berkas yang terjadi sepanjang waktu.

Manfaat menggunakan Git:

- Dokumentasi setiap perubahan versi dengan baik dan rapi
- Dapat membuat versi cabang / alternatif
- Dapat kembali ke checkpoint tertentu apabila tidak jadi menerapkan perubahan

```
$ git clone https://github.com/git-for-windows/git
Cloning into 'git'...
remote: Enumerating objects: 500937, done.
remote: Counting objects: 100% (3486/3486), done.
remote: Compressing objects: 100% (1415/1415), done.
remote: Total 500937 (delta 2494), reused 2917 (delta 2071), pack-reused 497451
Receiving objects: 100% (500937/500937), 221.14 MiB | 1.86 MiB/s, done.
Resolving deltas: 100% (362274/362274), done.
Updating files: 100% (4031/4031), done.
```



id/x partners



**Bagaimana penerapannya
di ID/X Partners?**



Thank You!



id/x partners