

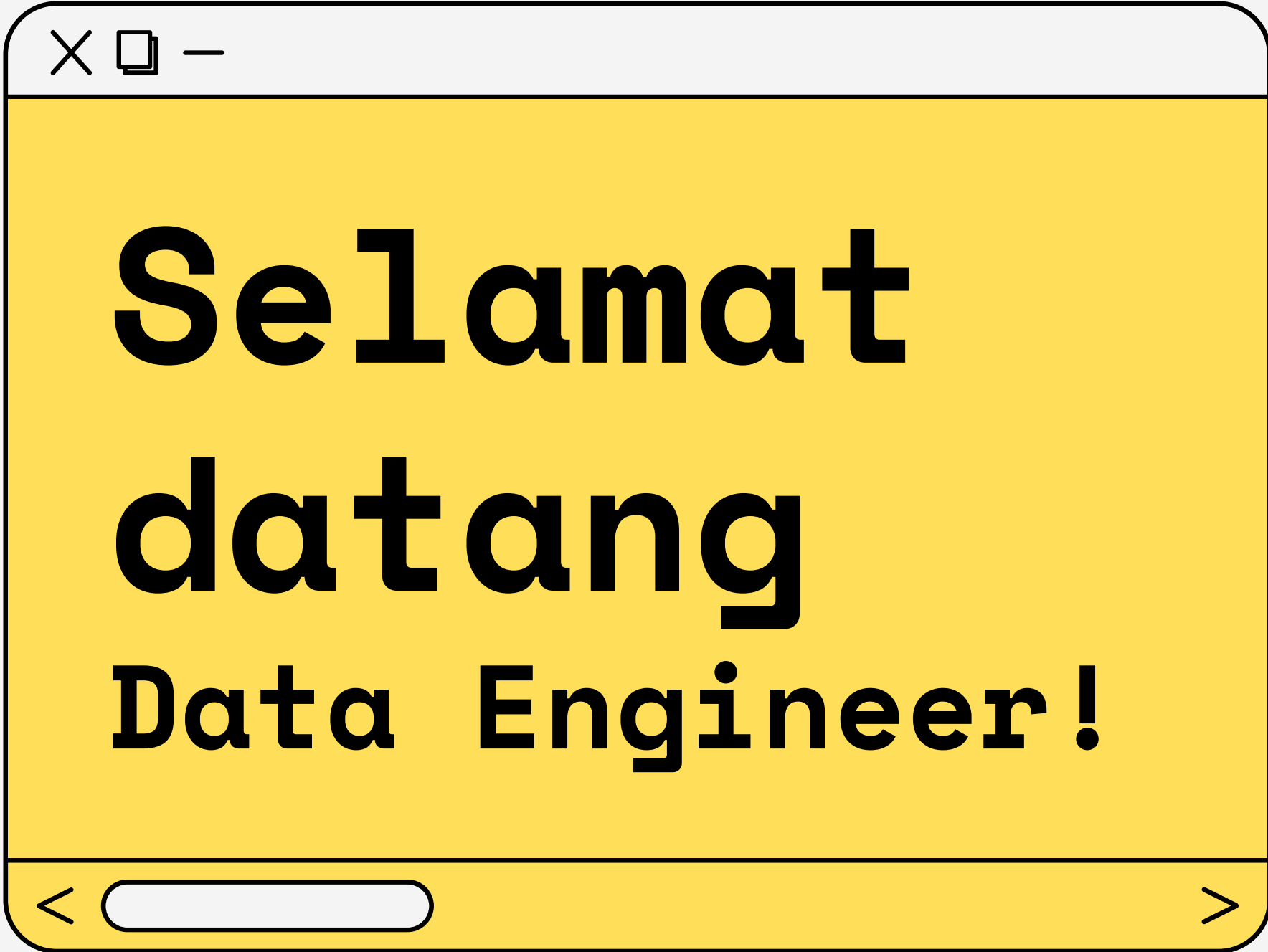
Data Engineer



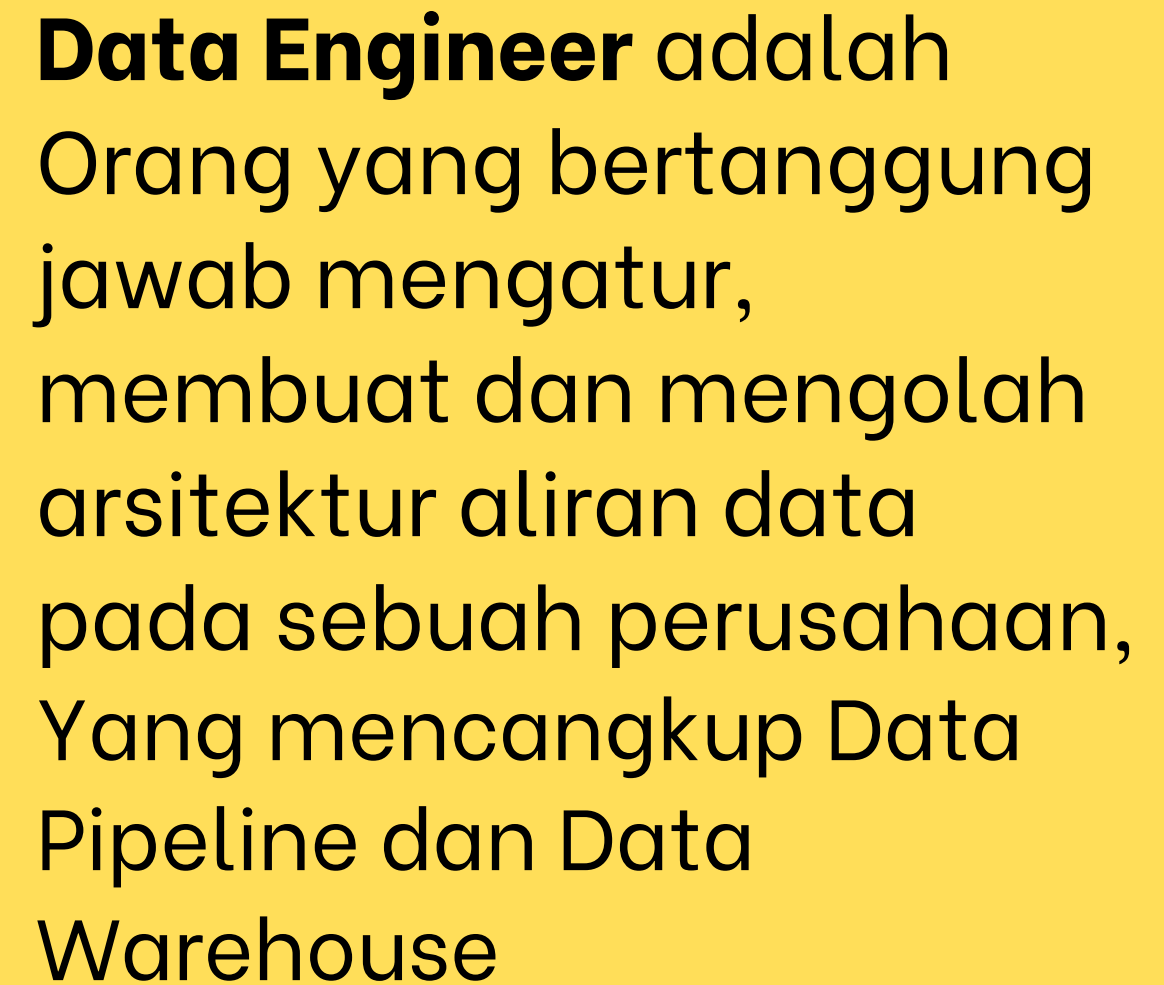
Learning Progress Review

by Charlie's Angels Team

- WEEK 1 -



Selamat datang Data Engineer!

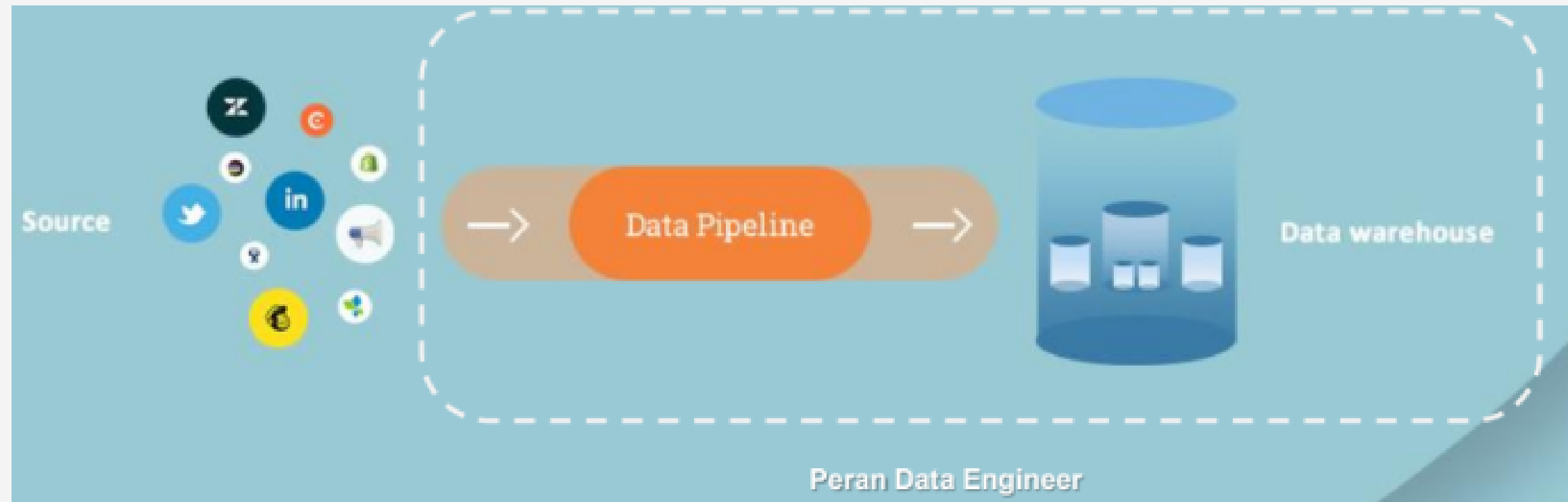


Data Engineer adalah
Orang yang bertanggung
jawab mengatur,
membuat dan mengolah
arsitektur aliran data
pada sebuah perusahaan,
Yang mencangkup Data
Pipeline dan Data
Warehouse



Apa itu Data Engineer ?

Work Flow Data



Data Source

Sumber Data yang akan diambil dari berbagai macam bentuk file.

EX: CSV, Excel, Pic, Word, dll

Pipeline

System distirbusi pengambilan data dari Data Source ke Data lake, Data Warehous atau Data Mart.

Data Warehouse

Penampungan keseluruhan data yang sudah menjadi satu format data.

Type Pengambilan Data

1

Streaming Processing

Data yang langsung diproses ke Datawarehouse/Analytics Dashboard

- + Lebih Mudah, Data selalu Update
- Biaya Mahal

2

ELT/ETL

Pengambilan data yang discedule, dan tidak bisa langsung

- + Murah
- Tidak bisa diambil setiap saat

Big Data

Big Data adalah kumpulan data yang sangat besar, kompleks dan terus bertambah setiap waktu. Data ini dihasilkan dari aktivitas internet yang makin rutin dilakukan, baik untuk tujuan pribadi maupun bisnis

Apa itu Big Data ?

Karakteristik Big Data

1

Volume

Karakteristik utama big data yaitu jumlah data yang sangat besar

2

Velocity

Kecepatan data juga menjadi karakteristik big data, meliputi input data hingga penggunaan data

3

Variety

Big data memiliki karakteristik yang bermacam-macam dan tidak homogen yaitu memiliki banyak sekali variabel dan sangat beragam

Karakteristik Big Data

4

Veracity

Big data bukan hanya tentang adanya data yang dihasilkan, tapi tentang identifikasi data dengan tepat agar memberikan manfaat bagi pengguna.

5

Value

Big data memiliki value untuk memudahkan pengguna mengakses informasi dengan cepat dan mengambil keputusan berdasarkan berbagai data yang ada.

Big Data Tools



1 SQL (Structured Query Language)

Bahasa pemrograman untuk mengambil, mengubah dan memanipulasi data dari sebuah database.



2 No SQL (Not Only SQL)

Konsep serta model basis data yang fleksibel dan tidak mengikuti kaidah-kaidah database relational



3 Airflow



4 Hadoop, Map Reduce, Spark, Hive



5 Docker



6 Kafka

Big Data Tools berbasis cloud



1 Google Cloud: Data proc, Cloud Composer



2 AWS: Redshift, DynamoDB



3 Alibaba Cloud: Alibaba MaxCompute



- Programming
- Python
- Data Types
- Variables
- Collections
- Modules, Packages,
Libraries
- Import



Programming

Secara sederhana, programming merupakan proses menulis program komputer untuk menjalankan suatu fungsi tertentu. Beberapa bahasa pemrograman yang populer digunakan antara lain : python, java, php, javascript, dan lain-lain.



Python






Python merupakan salah satu bahasa pemrograman yang saat ini sedang **populer** digunakan

- mudah dipelajari dan digunakan
- memiliki banyak library dan framework
- multifungsi

Why?



Versi python terbaru saat ini adalah python3. Namun versi sebelumnya yaitu python2 juga masih banyak digunakan

PYTHON 2.X		PYTHON 3.X
← LEGACY		FUTURE →
It is still entrenched in the software at certain companies		It will take over Python 2 by the end of 2019
 LIBRARY		LIBRARY 
Many older libraries built for Python 2 are not forwards compatible		Many of today's developers are creating libraries strictly for use with Python 3
0100 0001 ASCII		UNICODE 0000 0000 0100 0001
Strings are stored as ASCII by default		Text Strings are Unicode by default
≈ 7/2=3		7/2=3.5 =
It rounds your calculation down to the nearest whole number		This expression will result in the expected result
 print "WELCOME TO GEEKSFORGEEKS"		print("WELCOME TO GEEKSFORGEEKS") 

source : GeeksforGeeks

Data Types

Numeric : Integers (bilangan bulat), float (bilangan desimal)
Text : Strings, syntax diawali dan diakhiri dengan (“/”)
e.g : (“Hello”)
Boolean : True/false

Variables

- **Tempat untuk menyimpan data yang bersifat mutable**
 - Mutable : bersifat bisa dirubah (nilainya dapat diganti, ditambah, dihapus)
 - Immutable : bersifat tetap (nilainya tidak bisa diganti, ditambah, dihapus)
- **Aturan penulisan variables**
 - Diawali dengan huruf atau underscore (_), tidak boleh menggunakan angka
 - Case sensitive (huruf kapital/huruf kecil berpengaruh)
 - Tidak menggunakan keyword existing python (e.g : class)

Collections

- **List** : mutable, dapat menyimpan value yang sama
 - Syntax : `x = ["a", "b", "c"]`
- **Tuple** : immutable, dapat menyimpan nilai yang sama
 - Syntax : `x = ("a", "b", "c")`
- **Set** : mutable dan unique (tidak bisa menyimpan nilai yang sama)
 - Syntax : `x = {"a", "b", "c"}`
- **Dictionary** : mutable dan memiliki nilai yang berpasangan
 - Syntax : `x = {"nama": "Andi", "umur": "24"}`



Module

module pada python adalah ekstensi .py yang berisi skrip python. Module memiliki berbagai macam isi, baik itu fungsi, class, mau pun variabel.



Package

Sekumpulan modul python yang berada dalam sebuah folder dan memiliki satu modul constructor (`_init_.py`). File constructor berfungsi memberi tahu python interpreter bahwa folder tersebut adalah sebuah package.



Library

gabungan dari sekumpulan package dan modul dengan fungsionalitas yang sama dengan tujuan untuk memudahkan dalam membuat suatu aplikasi, tanpa harus menulis banyak kode.

Contoh library python 🔍



Library Python 🔍

Cara Install Library Python menggunakan PIP

PIP merupakan program untuk manajemen paket di Python. Tugasnya untuk menginstal, menghapus, upgrade paket Python, dan lain-lain.

```
pip install <nama_library>
```

Cara Import Library Python

setelah library terinstall dengan menggunakan pip, maka library telah siap untuk digunakan atau di import pada kode python. Gunakan perintah dibawah ini untuk import library python ke dalam kode /workspace.

```
import nama_library as alias
```




TERIMA KASIH!

Sampai jumpa di
Learning Progress Review
kami berikutnya!

▶ Hilda Meiranita Prastika Dewi

▶ Nur Indrasari

▶ Rezha Sulvian

▶ Tasha Dinya Ainsha