**Data Collecting Project**

**Studi Kasus: Twitter Data Crawling**

1. **Project**

Lakukan data collecting terhadap Twitter melalui dengan ketentuan sbb:

* Streaming dilakukan pada taggar **#BigData**.
* Jumlah data yang dicollect sebanyak 100 buah.
* Data yang dicollect adalah '**screen\_name**', '**lang**', '**created\_at**', '**retweet\_count**', '**text**', '**location**'
* Terus mencari data dengan retweet\_count dengan jumlah diatas 10
* Simpan dataframe hasil COLLECT ke dalam Amazon S3 dalam format CSV

1. **Project**

Lakukan data collecting terhadap Twitter melalui streaming dengan ketentuan sbb:

* Streaming dilakukan pada taggar **#BigData**.
* Jumlah data yang dicollect sebanyak 50 buah.
* Data yang dicollect adalah '**screen\_name**', '**lang**', '**created\_at**', '**retweet\_count**', '**text**', '**location**'
* Tampung data hasil streaming ke dalam list
* Konversikan data list hasil streaming ke dalam dataframe
* Simpan dataframe hasil streaming ke dalam Amazon S3 dalam format CSV
* Lakukan streaming (50 buah data) pada data Twitter dg keyword **#Senin**, data streaming disimpan di file **SeninTwitter.json** (Data stream yang disimpan adalah JSON utuh)

**Web Crawler Project**

**Studi Kasus: Website Berbasis Wordpress**

1. **Project**

Lakukan web crawler pada website berbasis wordpress dengan metode scrapping dengan ketentuan sbb:

* **Scrapping** hannya dilakukan pada website berbasis wordpress.
* Melakukan Scrapping dengan kedalaman atau *depth*  yang dapat ditentukan sendiri
* Data yang dicollect adalah '**judul**', '**deskripsi**', '**link**'
* Simpan hasil scrapping ke dalam format CSV.