

SOAL UJIAN TENGAH SEMESTER GENAP TA. 2021/2022

Mata Kuliah	:	Praktikum Big Data
Waktu	:	13:00 - Selesai
Hari/Tanggal	:	29 Maret 2022 – 2 April 2022
Program Studi	:	S-1 Informatika
Semester	:	6 (enam)
Dosen	:	Dr. Ermatita, S.Kom, MTI Oktavian Pratama, S.Kom, M.Kom I Wayan Widi Pradnyana, S.Kom, MTI
Sifat Ujian	:	Praktikum Individu Menggunakan Python Notebook

SOAL KASUS

Diketahui bahwa tahap Big Data Analytics Lifecycle yaitu : Business Case Evaluation, Data Identification, Data Acquisition & Filtering, Data Extraction, Data Validation & Cleansing, Data Aggregation & Representation, Data Analysis, Data Visualization, Utilization of Analysis Result.

Dari halaman berikut <http://data.kalbarprov.go.id/organization/dinas-pemberdayaan-masyarakat-dan-desa-prov-kalbar> , disediakan 3 data set :

- Daftar_Desa_dan_Kelurahan.json (<http://data.kalbarprov.go.id/dataset/e2f097e2-027c-4f6d-948d-ed85f94515c7/resource/56816d25-fa58-45db-b21d-664f8508573e/download/344c892f-fd4f-474b-83fc-806aaf6a6b80.json>)
- rekap-jumlah-status-desa-provinsi-kalimantan-barat-tahun-2019.csv (<http://data.kalbarprov.go.id/dataset/bfa647a7-fb24-4a9e-b3b4-12534cf558bf/resource/e739b4c3-38a7-42c3-b8d8-6c847ca45ef2/download/rekap-jumlah-status-desa-provinsi-kalimantan-barat-tahun-2019.csv>)
- LEMBAGA KEMASYARAKATAN DESA.pdf (<http://data.kalbarprov.go.id/dataset/58991af3-6f7f-4316-90fd-402b05ec22f8/resource/c39fd85b-e989-4ac0-89c8-a86527298e94/download/lembaga-kemasyarakatan-desa.pdf>)

Mahasiswa diminta melakukan penerapan Big Data Analytics Lifecycle pada lingkungan Big Data Computation Engine yaitu Apache Spark dengan bantuan Apache Spark terhadap beberapa data perkembangan desa suatu wilayah.

Buatlah sebuah Python Notebook yang berisikan beberapa cell untuk mengerjakan tahapan-tahapan soal berikut.

1. Business Case Evaluation. (10)
Berikan pemahaman mahasiswa mengenai potensi kasus bisnis yang mungkin menjadi permasalahan bisnis dari kombinasi 3 data tersebut. Definisikan menurut Anda : siapa pengguna data, kapan data dapat digunakan, mengapa menggunakan data tersebut,
2. Data Identification. (10)
Berikan pemahaman mahasiswa mengenai karakteristik 3 data tersebut, definisikan atribut datanya.
3. Data Acquisition & Filtering. (10)
Implementasikan tahapan pengambilan data dari sumber online data tersebut menuju penyimpanan lokal Spark. Poin ekstra jika mampu menyimpan dalam Apache Hive.
4. Data Extraction. (10)
Implementasikan proses pengambilan data dari json, csv, dan pdf menjadi Data Frame Spark atau RDD Spark.
5. Data Validation & Cleansing. (10)
Implementasikan proses perapihan yang diperlukan terhadap masing-masing data.
6. Data Aggregation & Representation. (10)

Implementasikan proses kombinasi antara data set tersebut, dan lakukan statistik dasar.

7. Data Analysis. (20)

Implementasikan proses analisis yang diperlukan, dapat berupa eksploratory, classification, clustering, dsb.

8. Data Visualization. (10)

Implementasikan 2 buah visualisasi yang menurut Anda dapat memberikan inspirasi untuk pengguna data ini.

9. Utilization of Analysis Result. (10)

Berikan saran Anda bagaimana potensi pemanfaatan sarana big analytics ini, apakah Input for Enterprise System, Business Process Optimization, atau Alers. Berikan penjelasan Anda.

“Selamat Mengerjakan”

VALIDASI SOAL

Jakarta,
Mengetahui & Menyetujui
Ka. Program Studi S-1 Informatika

Tim Dosen MK Big Data
Koordinator

(Desta Sandya Prasvita, S.Kom, M.Kom)

(Dr. Ermatita, S.Kom, MTI)