

## SOAL UJIAN AKHIR SEMESTER GENAP 2021-2022

**Telah diverifikasi:**

27 Mei 2022

**Dr. Dedi Trisnawarman**  
Kaprodi S

Program Studi : Sistem Informasi

Kelas : A, B dan S

Mata Kuliah : Big Data

Tanggal : 02 Juni 2022

Dosen : Ir. Jeanny Pragantha, M.Eng  
Teny Handhayani, Ph.D  
Tony, Ph.D  
Dra. Chairisni Lubis, M.Kom

Waktu : 07:30-09:30 (120 menit) (A,B)  
17:00-19:00 (120 menit) (S)

Kode MK : SI23016

Sifat : Open book pada Ms. Team  
Ujian melalui menu Quiz

Catatan : -

*Bila dalam pelaksanaan ujian peserta melanggar tata tertib ujian dan berlaku curang, maka nilai ujian langsung diberi nilai E tanpa memperhatikan komponen nilai lain.*

### Capaian Pembelajaran (CPL) dan Capaian Pembelajaran Mata Kuliah (CPMK) \*\*

	CPL				
	S8	P2	KU1	KU5	KK17
CPMK 01	√			√	
CPMK 02		√			
CPMK 03					√
CPMK 04		√			√

#### Keterangan:

#### Capaian Pembelajaran (CPL): \*

- S8** Menunjukkan sikap bertanggungjawab atas pekerjaan di bidang keahliannya secara mandiri.
- P2** Menguasai konsep dan ilmu probabilitas dan statistik untuk mendukung dan menganalisis data dan informasi.
- KU1** Mampu menerapkan pemikiran logis, kritis, sistematis, dan inovatif dalam konteks pengembangan atau implementasi ilmu pengetahuan dan teknologi yang memperhatikan dan menerapkan nilai humaniora yang sesuai dengan bidang keahliannya.
- KU5** Mampu mengambil keputusan secara tepat dalam konteks penyelesaian masalah di bidang keahliannya berdasarkan hasil analisis informasi dan data.
- KK17** Mampu memahami dan mengikuti perkembangan teknologi Big Data.

**Capaian Pembelajaran Mata Kuliah:\***

**CPMK 01** Mahasiswa memahami konsep dasar Big Data. (S8,KU5)

**CPMK 02** Mahasiswa mampu memahami teknik analitik terhadap Big Data (P2)

**CPMK 03** Mahasiswa mengenal dan dapat mengikuti perkembangan teknologi Big Data. (KK17)

**CPMK 04** Mahasiswa mampu menggunakan teknologi Big Data.(P2,KK17)

No.	Bentuk Soal (PG/Essay /Kasus)	URAIAN MATERI	CPMK**				BOBOT SOAL (%)
			CPMK 01	CPMK 02	CPMK 03	CPMK 04	
1.	PG	Teknologi penyimpanan	√		√		1.8
2.	PG	RDBMS database	√		√		1.8
3.	PG	RDBMS database	√		√		1.8
4.	PG	RDBMS database	√		√		1.8
5.	PG	RDBMS database	√		√		1.8
6.	PG	NoSQL database	√		√		1.8
7.	PG	NoSQL database	√		√		1.8
8.	PG	NoSQL database	√		√		1.8
9.	PG	NoSQL database	√		√		1.8
10.	PG	NoSQL database	√		√		1.8
11.	PG	NoSQL database	√		√		1.8
12.	PG	NoSQL database	√		√		1.8
13.	PG	NoSQL database	√		√		1.8
14.	PG	NoSQL database	√		√		1.8
15.	PG	NoSQL database	√		√		1.8
16.	PG	<i>In memory storage devices</i>	√		√		1.8

No.	Bentuk Soal (PG/Essay /Kasus)	URAIAN MATERI	CPMK**				BOBOT SOAL (%)
			CPMK 01	CPMK 02	CPMK 03	CPMK 04	
17.	PG	<i>On disk storage devices</i>	√		√		1.8
18.	PG	Perangkat penyimpanan NoSQL	√		√		1.8
19.	PG	Perangkat penyimpanan NoSQL	√		√		1.8
20.	PG	Perangkat penyimpanan NoSQL	√		√		1.8
21.	PG	Analisis kuantitatif		√			1.8
22.	PG	Analisis kualitatif		√			1.8
23.	PG	Data mining		√			1.8
24.	PG	Analisis statistik		√			1.8
25.	PG	Correlation		√			1.8
26.	PG	A/B testing		√			1.8
27.	PG	Pemahaman analisis statistik		√			1.8
28.	PG	Analisis regresi		√			1.8
29.	PG	Implementasi regresi		√			1.8
30.	PG	Klasifikasi		√			1.8
31.	PG	Clustering		√			1.8
32.	PG	Visualisasi		√			1.8
33.	PG	Filtering		√			1.8
34.	PG	Analisis sentimen		√			1.8
35.	PG	Natural Language Processing (NLP)		√			1.8
36.	PG	Analisis visual		√			1.8

No.	Bentuk Soal (PG/Essay /Kasus)	URAIAN MATERI	CPMK**				BOBOT SOAL (%)
			CPMK 01	CPMK 02	CPMK 03	CPMK 04	
37.	PG	Time series		✓			1.8
38.	PG	Analisis jaringan		✓			1.8
39.	PG	Data spasial		✓			1.8
40.	PG	Manfaat analisis jaringan		✓			1.8
41.	Essay	Teknik analysis Big Data		✓			14
42.	Essay	Tutorial Python		✓		✓	14

\* : CPL dan CPMK disesuaikan dengan yang tertulis di RPS

\*\* : beri tanda ✓

\*\*\* : materi pembelajaran yang terkait soal ujian

**PETUNJUK MENGERJAKAN SOAL:**

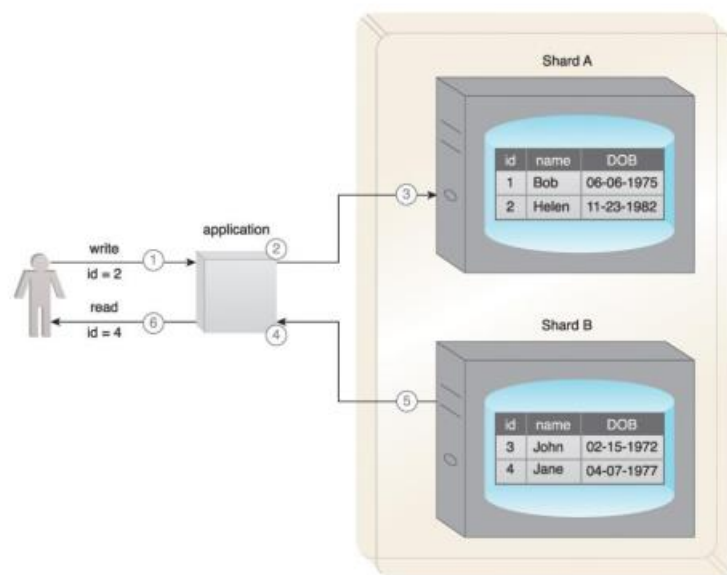
1. Jawablah semua soal-soal berikut dengan benar!
2. Bagi yang terlambat akan dikenakan sanksi berupa pengurangan nilai.  
Pengurangan nilai juga akan dilakukan jika jawaban mahasiswa tidak memenuhi ketentuan di atas atau berlaku curang.

**# Pilihan Ganda (nilai = 72 @1.8)**

1. Istilah untuk sistem file yang tidak mengenal atau mengetahui data yang disimpan adalah ...

A	diagnostic
B	irrelevant
C	agnostic
D	dedicated
E	centralized

Gambar berikut untuk soal no. 2 dan 3:



2. Apa yang terjadi pada ②?

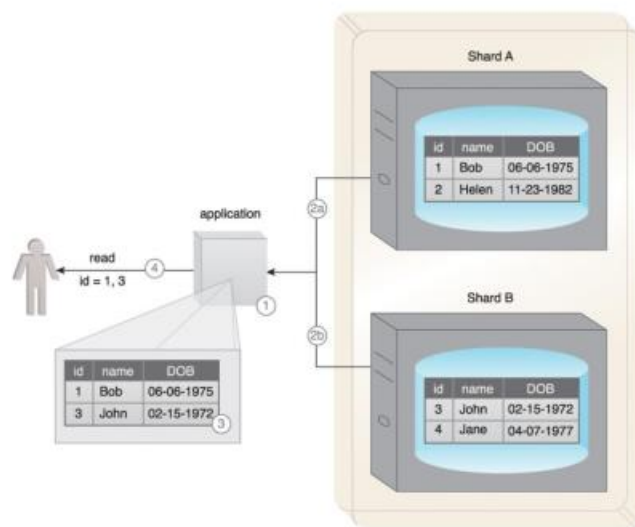
A	<i>user</i> menulis sebuah <i>record</i>
B	data dibaca dan dikembalikan ke aplikasi
C	aplikasi mengembalikan <i>record</i> ke <i>user</i>

D	aplikasi menentukan <i>shard</i> mana yang harus ditulis
E	<i>record</i> dikirim ke <i>shard</i> yang ditentukan aplikasi

3. Apa yang dilakukan di ⑤ ?

A	<i>user</i> menulis sebuah <i>record</i>
B	data dibaca dan dikembalikan ke aplikasi
C	aplikasi mengembalikan <i>record</i> ke <i>user</i>
D	aplikasi menentukan <i>shard</i> mana yang harus ditulis
E	<i>record</i> dikirim ke <i>shard</i> yang ditentukan aplikasi

Perhatikan gambar berikut untuk menjawab soal nomor 4 dan 5!



4. ① menunjukkan ....

A	<i>user</i> meminta <i>record</i> (id = 1,3) dan aplikasi menentukan <i>shard</i> yang perlu dibaca
B	data dikembalikan ke <i>user</i>
C	aplikasi mengembalikan <i>record</i> ke <i>user</i>
D	aplikasi menentukan <i>shard</i> mana yang harus ditulis
E	aplikasi menentukan bahwa <i>shard</i> A dan <i>shard</i> B perlu dibaca

5. ④ menunjukkan ....

A	<i>user</i> meminta <i>record</i> (id = 1,3) dan aplikasi menentukan <i>shard</i> yang perlu dibaca
B	data dikembalikan ke <i>user</i>
C	aplikasi mengembalikan <i>record</i> ke <i>user</i>
D	aplikasi menentukan <i>shard</i> mana yang harus ditulis
E	aplikasi menentukan bahwa <i>shard</i> A dan <i>shard</i> B perlu dibaca

6. Berikut ini yang termasuk karakteristik NoSQL adalah .....

A	<i>scale out</i>
B	<i>highly available</i>
C	<i>high speed</i>
D	<i>lower operational cost</i>
E	<i>eventual consistency</i>

7. Karakteristik NoSQL yang dibangun di atas platform *open source* tanpa biaya lisensi adalah .....

A	<i>scale out</i>
B	<i>highly available</i>
C	<i>high speed</i>
D	<i>lower operational cost</i>
E	<i>eventual consistency</i>

8. Karakteristik NoSQL yang menghilangkan kebutuhan akan lapisan caching terdistribusi pihak ketiga disebut .....

A	<i>scale out</i>
B	<i>highly available</i>
C	<i>integrated caching</i>
D	<i>lower operational cost</i>
E	<i>eventual consistency</i>

9. Salah satu karakteristik NoSQL adalah *API driven data access*, yang artinya .....

A	menghilangkan kebutuhan akan lapisan <i>caching</i>
B	mendukung penskalaan horizontal
C	data memiliki bentuk mentah
D	mendukung kueri berbasis API

E	tidak memiliki biaya lisensi
---	------------------------------

10. Ada 4 (empat) jenis perangkat penyimpanan NoSQL, di antaranya .....

A	<i>graph</i>
B	<i>key-chain</i>
C	<i>document</i>
D	<i>column-family</i>
E	<i>key-value</i>

11. Gambar berikut menunjukkan jenis NoSQL *storage* .....



A	<i>graph</i>
B	<i>key-chain</i>
C	<i>document</i>
D	<i>column-family</i>
E	<i>key-value</i>

12. Gambar berikut menggambarkan jenis NoSQL *storage* .....

key	value
631	John Smith, 10.0.30.25, Good customer service
365	10010101110110111101110101010101001110011010
198	<CustomerId>32195</CustomerId><Total>43.25</Total>

A	<i>graph</i>
B	<i>key-chain</i>
C	<i>document</i>
D	<i>column-family</i>



E	<i>key-value</i>
---	------------------

13. Berikut ini yang termasuk perangkat penyimpanan dokumen adalah .....

A	Redis
B	MongoDB
C	Amazon DynamoDB
D	CouchDB
E	Terrastore

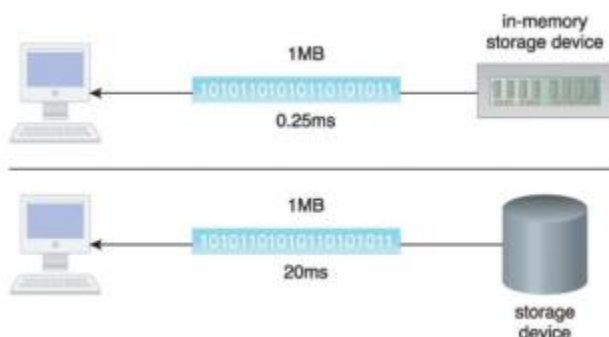
14. Cassandra termasuk perangkat penyimpanan .....

A	<i>graph</i>
B	<i>key-chain</i>
C	<i>document</i>
D	<i>column-family</i>
E	<i>key-value</i>

15. Perangkat penyimpanan dalam memori (*in-memory*) dapat diimplementasikan sebagai .....

A	<i>in-memory data base</i>
B	<i>in-memory data set</i>
C	<i>in-memory data track</i>
D	<i>in-memory data trip</i>
E	<i>in-memory data grid</i>

16. Perhatikan gambar berikut!



Penyimpanan *in-memory* lebih cepat diperkirakan ..... kali lebih cepat dari pada penyimpanan disk.

A	20
B	40
C	60
D	80
E	100

17. *On-disk storage disk* menggunakan *hard disk drive* berbiaya rendah untuk penyimpanan jangka panjang.

A	Benar
B	Salah

18. BLOB adalah singkatan untuk *Big Large Object*.

A	Benar
B	Salah

19. Dokumen semi-terstruktur terdiri dari *flat-schema* dan *nested-schema*.

A	Benar
B	Salah

20. *Key-value NoSQL storage* mendukung *partial updates*.

A	Benar
B	Salah

21. Berikut pernyataan benar mengenai analisis kuantitatif yaitu ....

A	menggunakan sampel yang lebih kecil
B	berfokus pada kuantifikasi pola dan korelasi dalam data
C	melibatkan sejumlah besar data
D	hasil analisis tidak dapat diukur secara numerik
E	hasil analisis dapat digunakan untuk perbandingan numerik

22. Berikut pernyataan benar mengenai analisis kualitatif yaitu ....

A	hasil analisis tidak dapat digeneralisasikan ke seluruh kumpulan
B	melibatkan sejumlah besar data
C	hasil analisis tidak dapat diukur secara numerik
D	hasil analisis dapat digunakan untuk perbandingan numerik
E	menggunakan sampel yang lebih kecil

23. Berikut pernyataan benar mengenai data mining yaitu ....

A	memiliki tujuan mengidentifikasi pola yang sebelumnya tidak diketahui
B	digunakan untuk analitik prediktif
C	menggunakan kumpulan data besar
D	menggunakan sampel berukuran kecil
E	menggunakan teknik berbasis perangkat lunak otomatis

24. Analisis statistik selalu bersifat kuantitatif.

A	Benar
B	Salah

25. Correlation membandingkan dua versi elemen untuk menentukan versi mana yang lebih unggul berdasarkan metrik yang telah ditentukan sebelumnya.

A	Benar
B	Salah

26. Berikut pernyataan benar mengenai A/B testing yaitu ....

A	melibatkan elemen <i>control</i> dan <i>treatment</i>
B	melibatkan data sampel yang berukuran kecil
C	kedua elemen mengalami percobaan secara bersamaan
D	untuk membandingkan dua produk
E	percobaan tidak dilakukan bersamaan

27. Apa jenis teknik analisis yang tepat untuk mengetahui efektivitas beberapa jenis vaksin untuk yaitu ....

A	Correlation
---	-------------

B	Regression
C	A/B Testing
D	Clustering
E	Classification

28. Berikut pernyataan benar mengenai teknik analisis regresi yaitu ....

A	menganalisis keterkaitan variabel dependen dan variabel independen
B	hasil analisis tidak dapat diukur secara numerik
C	melibatkan elemen <i>control</i> dan <i>treatment</i>
D	beberapa variabel independen dapat diuji secara bersamaan
E	menggunakan teknik berbasis perangkat lunak otomatis

29. Teknik yang tepat untuk memprediksi harga rumah berdasarkan luas area yaitu ....

A	Correlation
B	Regression
C	A/B Testing
D	Clustering
E	Classification

30. Berikut pernyataan benar tentang klasifikasi yaitu ....

A	Supervised learning
B	Unsupervised learning
C	kategori secara implisit dihasilkan berdasarkan pengelompokan data
D	Klasifikasi dapat dilakukan untuk dua atau lebih kategori
E	menganalisis keterkaitan variabel dependen dan variabel independen

31. Berikut pernyataan benar tentang clustering yaitu ....

A	Supervised learning
B	Unsupervised learning
C	kategori secara implisit dihasilkan berdasarkan pengelompokan data
D	melibatkan elemen <i>control</i> dan <i>treatment</i>
E	Representasi visual biasanya menggunakan scatter plot

32. Grafik boxplot dapat digunakan untuk visualisasi data outlier.

A	Benar
---	-------

B	Salah
---	-------

33. Collaborative filtering adalah teknik pemfilteran item berdasarkan kolaborasi, atau penggabungan, dari perilaku masa lalu pengguna dengan perilaku orang lain.

A	Benar
B	Salah

34. Analisis sentiment merupakan bagian dari analisis semantik.

A	Benar
B	Salah

35. Berikut pernyataan benar mengenai Natural Language Processing (NLP) yaitu ....

A	digunakan untuk pengenalan teks dan ucapan
B	digunakan untuk mencari kelompok data
C	Supervised learning
D	Unsupervised learning
E	termasuk dalam jenis analisis semantik

36. Teknik analisis visual yang efektif untuk mengekspresikan pola, komposisi data melalui hubungan sebagian-keseluruhan dan distribusi geografis data yaitu ....

A	Heat map
B	Scatter plot
C	Box plot
D	Pie chart
E	Bar chart

37. Analisis yang melibatkan kumpulan nilai berurutan yang direkam dalam interval waktu yang teratur yaitu ....

A	Natural Language Processing
B	Text Analytics
C	Collaboratif filtering
D	Sentiment Analysis
E	Time series

38. Teknik yang berfokus menganalisis hubungan antara entitas dalam jaringan yaitu ....

A	Network graph
B	Text Analytics
C	Heat map
D	Sentiment Analysis
E	Time series

39. Analisis data spasial difokuskan pada analisis data berbasis lokasi untuk menemukan hubungan dan pola geografis yang berbeda antar entitas.

A	Benar
B	Salah

40. Manfaat analisis jaringan yaitu ....

A	Optimasi rute
B	Prediksi penyebaran penyakit
C	Analysis jaringan sosial
D	Prediksi cuaca
E	Prediksi harga rumah

## ## ESSAY (nilai = 28 @ 14)

Berikut ini merupakan sampel data pasien Covid-19 yang bertahan hidup dan meninggal yang ditunjukkan oleh variabel survival (hidup = 0, meninggal = 1).

Asumsikan bahwa Anda memiliki 10000 sampel data dengan jumlah pasien hidup 8000 orang dan meninggal 2000 orang. Anda diminta untuk melakukan analisis pada data berikut!

Detak jantung	SpO2	Pasang_O2	Temperatur	Respiratory Rate	Survival
120	97.711213	0	37	20	0
3	99	0	36.3	20	0
102	99	0	36.3	20	0
1	100	0	36	20	0
88	100	0	36	20	0
1	97.711213	0	36	20	0
89	97.711213	0	36	20	0
1	97	0	37	22	0
110	99	0	36.1	23	0
106	97.711213	0	37.5	20	0
120	97.711213	0	36.5	20	0
110	95.665794	0	36.4	22	1
90	100	0	36.5	20	1
90	100	0	36.5	20	1
122	95.665794	0	37	22	1
140	92	0	37.6	20	1

41. Sebutkan dan jelaskan teknik analisis Big Data yang Anda gunakan!

42. Tulis Python code untuk:

- Load data set tersebut dengan menggunakan Pandas. Lokasi penyimpanan file csv ditentukan sendiri.
- Tampilkan 5 (lima) record terakhir dari data set tersebut.
- Tampilkan jumlah baris dan jumlah kolom dari data set tersebut.
- Tampilkan data pasien yang memiliki temperature tertinggi dan terendah.
- Tampilkan data pasien dengan temperatur di atas 36 dan respiratory rate di atas 20.
- Tampilkan data 20 pasien dengan nilai rata-rata SpO2 tertinggi.
- Ubah nama kolom pada data set tersebut menjadi **detak\_jantung**, **nilai\_SpO2**, **O2\_pasang**, **nilai\_temperatur**, **respiratory\_rate**, dan **survival\_rate**.

=====SELAMAT MENGERJAKAN=====