

**UJIAN AKHIR SEMESTER
BIG DATA
DOSEN : Imam Haromain, S.Si., M.Kom.**

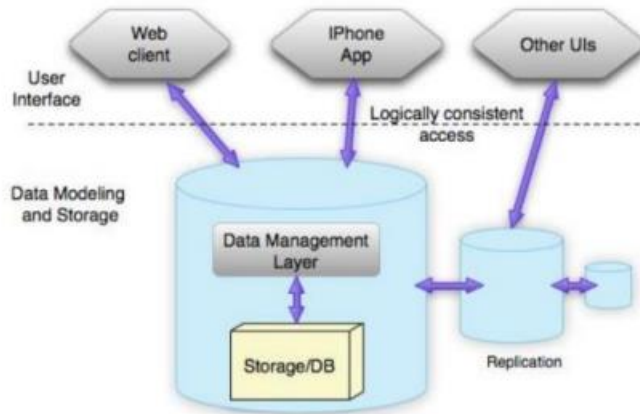
**Alfian Nur Usyaid
(0110222132)**



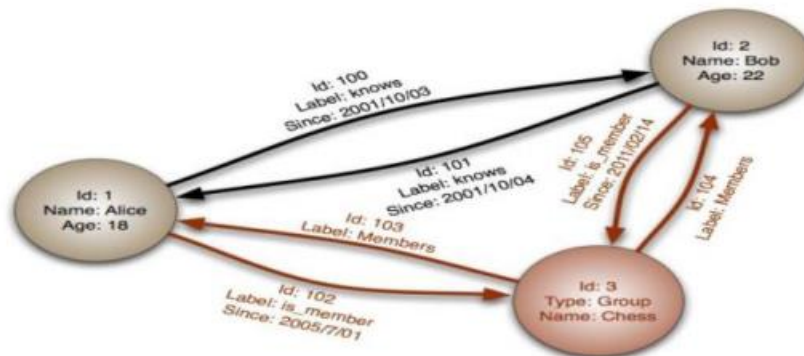
**STT TERPADU NURUL FIKRI
PROGRAM STUDI TEKNIK INFORMATIKA
DEPOK
2024**

SOAL :

1. (5%) Jelaskan apa yang dimaksud dengan Teorema CAP pada sistem basis data terdistribusi ? Dimana posisi MongoDB pada Teorema CAP ? (CPMK02)
2. (5%) Buatkan kueri pada basis data MongoDB dengan kondisi kueri yang melewati 5 pelanggan pertama dan mengembalikan 10 pelanggan berikutnya ? (CPMK02)
3. (5%) Pada proses Cross-Industry Standard Process for Data Mining atau CRISP-DM ada fase data understanding dan data preparation, apa beda fase tersebut, jelaskan ? (CPMK02)
4. (5%) Apakah Elasticsearch menyediakan antarmuka pengguna grafis interaktif untuk para penggunanya ? (CPMK02)
5. (20%) Jelaskan ilustrasi dari gambar dibawah ini ? (CPMK02)



6. (20%) Jelaskan ilustrasi dari gambar dibawah ini terkait Basis Data NoSQL - Graph? (CPMK02)



7. Ikuti langkah langkah berikut : Unduh file account.json dari https://drive.google.com/file/d/1EQK25JdiV_5c1VPotnQv91_rIdFJ5Mg/view?usp=sharing (file accounts.json). (CPMK02)

- a. (5%) Masukkan konten tersebut ke server elasticsearch dengan nama index nya 2 nama depan kalian (contoh : imamharomain, jika nama hanya ada 1 kata tidak papa, contoh : imam). Screenshoot prosesnya, boleh melalui postman atau yang lainnya.
 - b. (5%) Cari data dengan account balance-nya lebih kecil sama dengan dari 4 angka terakhir nim anda.
 - c. (10%) Cari data orang yang memiliki nama depan sama dengan 2 huruf nama anda gunakan fungsi prefix.
8. Diberikan sebuah kasus dalam dunia kesehatan untuk mendeteksi penyebaran penyakit tertentu dengan memanfaatkan Big Data. (CPMK02)
- a. (5%) Data apa saja yang perlu dikumpulkan?
 - b. (5%) Informasi/insight apa saja yang bisa didapatkan dari data tersebut?
 - c. (10%) Teknologi big data apa yang akan anda gunakan dalam mendapatkan data dan informasi tersebut, mengapa anda memilih teknologi tersebut?

Jawaban

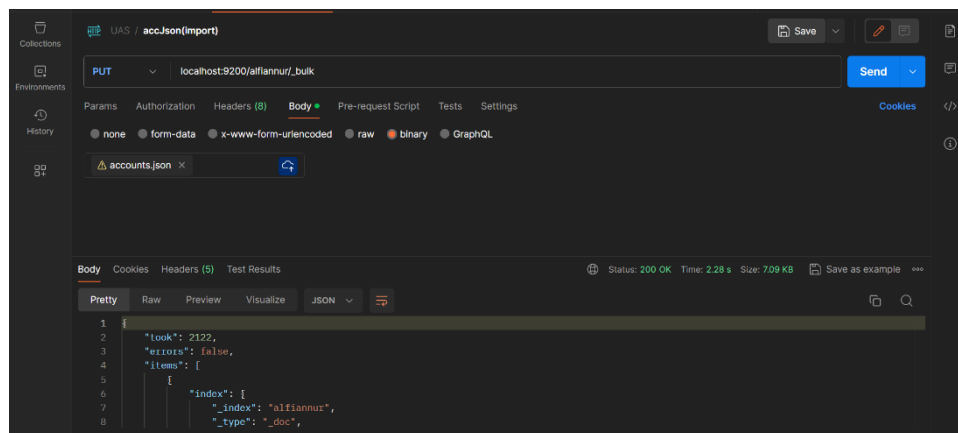
1. Teorema CAP adalah Teori di mana setiap Sistem Basis Data yang terdistribusi dapat memberikan 3 Jaminan yaitu Consistency, Availability dan Partition Tolerance. Dan posisi MongoDB ada di antara Consistency dan Partition Tolerance (CP)
- 2.

```
mongosh mongodb://127.0.0.1:27021/
UAS_BigData> db.pelangganAlfian.find().skip(5).limit(10).pretty();
[
  {
    _id: ObjectId("65a8858fa21fa1e73a4815fa"),
    nama: 'LIand',
    jk: 'Laki-laki'
  },
  {
    _id: ObjectId("65a8859ba21fa1e73a4815fb"),
    nama: 'Ka Nurmila',
    jk: 'Perempuan'
  },
  {
    _id: ObjectId("65a885a6a21fa1e73a4815fc"),
    nama: 'Jasmine',
    jk: 'Perempuan'
  },
  {
    _id: ObjectId("65a885b6a21fa1e73a4815fd"),
    nama: 'Kevin',
    jk: 'Laki-laki'
  },
  {
    _id: ObjectId("65a885cda21fa1e73a4815fe"),
    nama: 'Diablo',
    jk: 'Laki-laki'
  },
  {
    _id: ObjectId("65a88605a21fa1e73a4815ff"),
    nama: 'Belzebuub',
    jk: 'Laki-laki'
  },
  {
    _id: ObjectId("65a8860aa21fa1e73a481600"),
    nama: 'Muhammad',
    jk: 'Laki-laki'
  },
  {
    _id: ObjectId("65a88616a21fa1e73a481601"),
    nama: 'Aisyah',
    jk: 'Perempuan'
  },
  {
    _id: ObjectId("65a88627a21fa1e73a481602"),
    nama: 'Ragnarok',
    jk: '-'
  },
  {
    _id: ObjectId("65a88639a21fa1e73a481603"),
    nama: 'macem-macem',
    jk: 'dll'
  }
]
UAS_BigData>
```

3. Fase Data Understanding itu adalah fase untuk memeriksa data sehingga dapat mengidentifikasi masalah pada data. Sedangkan Fase Data Preparation adalah fase untuk memperbaiki masalah dalam data yang sudah diidentifikasi di Fase Data Understanding.
4. Iyaa, misalnya dengan menggunakan bantuan Kibana.
5. Gambar tersebut menunjukkan diagram arsitektur data modeling dan penyimpanan. Diagram ini terdiri dari dua lapisan utama, yaitu lapisan antarmuka pengguna (UI) dan lapisan manajemen data. Dalam gambar tersebut, ditunjukkan bahwa lapisan antarmuka pengguna dapat mengakses data secara logis dan konsisten. Artinya, data yang ditampilkan oleh lapisan antarmuka pengguna akan selalu konsisten, terlepas dari sumber datanya. Yang Intinya User Interface mengirimkan dan menerima Data melalui *Data Modeling and Storage* Begitupun sebaliknya dengan *Data Management Layer* juga bisa mengirimkan dan menerima Data untuk disimpan di *Database* yang siap untuk dikirimkan ke User Interface yang berhubungan juga dengan replikanya .
6. Gambar tersebut menunjukkan ilustrasi dari basis data NoSQL - Graph. Di gambar tersebut, terdapat tiga node, yang masing-masing memiliki name Alice, Bob dan Chess Club, khusus Chess Club sendiri memiliki *type* Group yang mana Alice dan Bob memiliki hubungan member dengan Chess Club melalui Edge label 'Members' sedangkan hubungan antara Alice dan Bob adalah sekedar "Tahu" atau "*Knows*" diwakili oleh edge dengan label "knows".

7. Menggunakan Postman

a.



b.

The screenshot shows a Postman interface with a GET request to `localhost:9200/alfiannur/_search`. The request body contains a GraphQL query:

```
1 {
2   ... "query" : {
3     ... "range" : {
4       ... "balance" : {
5         ... "lte" : 2132
6       }
7     }
8   }
9 }
```

The response is a JSON object with the following structure:

```
{
  "_source": {
    "account_number": 87,
    "balance": 1133,
    "firstname": "Hewitt",
    "lastname": "Kidd",
    "age": 22,
    "gender": "M",
    "address": "446 Halleck Street",
  }
}
```

Status: 200 OK, Time: 59 ms, Size: 1.43 KB

c.

The screenshot shows a Postman interface with a GET request to `localhost:9200/alfiannur/_search`. The request body contains a GraphQL query:

```
1 {
2   ... "query" : {
3     ... "prefix" : {
4       ... "firstname" : {
5         ... "value" : "al"
6       }
7     }
8   }
9 }
```

The response is a JSON object with the following structure:

```
{
  "_source": {
    "account_number": 436,
    "balance": 27585,
    "firstname": "Alexander",
    "lastname": "Sargent",
    "age": 23,
    "gender": "M",
    "address": "363 Albemarle Road",
  }
}
```

Status: 200 OK, Time: 132 ms, Size: 1.43 KB

8.

- a. Data-data yang perlu dikumpulkan adalah kita harus mengidentifikasi gen penyakit dan biomarker, catatan dokter, data perjalanan pasien, data jalur dan grafik serta data penagihan dan klaim.
- b. Dengan mengumpulkan Informasi data gen penyakit dan biomarker kita jadi bisa membantu menentukan masalah Kesehatan yang mungkin mereka hadapi di masa depan. Selain itu kita juga perlu mengumpulkan data pengihan dan klaim untuk menemukan peluang pendapatan yang hilang dan tempat-tempat di mana arus kas pembayaran dapat ditingkatkan.
- c. Teknologi yang mungkin saya gunakan adalah Cassandra karena tantangan untuk mendapatkan data di atas adalah volume data genom sangat besar, dan menjalankan algoritma yang rumit pada data tersebut, selain itu data tersebut membutuhkan waktu pemrosesan yang lama dalam jumlah besar yang beberapa di antaranya adalah multi-struktur. Sebab itulah Saya memilih Teknologi Cassandra yang memiliki jaminan Availability dan Partition Tolerance (AP) di Teorema CAP.