

Nama : Muhammad Azhar Rasyad
NIM : 0110217029
Program Studi : Teknik Informatika

Ujian Akhir Semester Big Data

1. (15%) Jelaskan apa yang dimaksud dengan CAP Teorema Pada sistem database terdistribusi ? Dimana posisi MongoDB pada CAP Teorema?

Jawaban : CAP atau Consistency Availability Partition Tolerance Theorem merupakan teori yang digunakan sebagai tool untuk membuat perancangan sistem agar aware akan trade-off saat merancang networked shared-data system, akan tetapi keterpenuhan CAP hanya 2 yang mungkin terpenuhi yaitu CA, CP, atau AP.

Posisi MongoDB pada CAP Teorema yaitu berada pada Consistency dan Partition Tolerance, karena MongoDB mempunyai sifat bahwa client selalu memiliki tampilan yang sama pada datanya (Consistency) dan sifat sistem berfungsi dengan baik di seluruh partisi pada jaringan fisik (Partition Tolerance), akan tetapi tidak memiliki Availability.

2. (10%) Sebutkan 4 model data No-SQL dan jelaskan masing-masing karakteristiknya dan contoh software aplikasinya?

Jawaban : Ada 4 model data No-SQL yaitu :

- Graph Database, memiliki karakteristik sebagai berikut :
 - Graph database diambil berdasarkan Teori Graph
 - Databasenya didesain berdasarkan relasi yang mewakili sebuah graph dan elemen interkoneksi graph
 - Data tersimpan didalam node dan edge
 - Setiap node diorganisasikan berdasarkan relasi antar node
 - Setiap node dan relasi yang terjadi terdapat pendefinisian properti
 - Contoh software aplikasinya yaitu Neo4j dan Titan
- Key-Value Store, memiliki karakteristik sebagai berikut :

- Databasenya didesain untuk menyimpan data dalam bentuk format minimum skema
- Semua data tersimpan dalam format berpasangan key–value, yang didalamnya terdapat proses index key dan value
- Contoh software aplikasinya yaitu Cassandra, DyanmoDB, Azure Table Storage, Riak, BerkeleyDB, dan Redis
- Column Store, memiliki karakteristik sebagai berikut :
 - Sebuah key dapat diperluas menyimpan data bentuk kolom dengan banyak dimensi
 - Pendekatan penyimpanan data table sebagai bagian dari column data ketimbang sebagai row data
 - Sebagai inverse dari bentuk standard database (RDBMS)
 - Contoh software aplikasinya yaitu Hbase, Big Table, dan HyperTable
- Document Database, memiliki karakteristik sebagai berikut :
 - Pengembangan dari Key–Value Store yang disimpan dalam bentuk sebuah dokumen yang lebih kompleks
 - Setiap dokumen memiliki unique key yang akan digunakan sebagai kunci untuk mengambil dokumen
 - Database didesain untuk penyimpanan, pengambilan, dan pengelolaan dengan format document-oriented information yang dikenal dengan semi structured data
 - Memiliki struktur data berformat seperti XML, JSON, BSON
 - Contoh software aplikasinya yaitu MongoDB dan CouchDB

3. (10%) Sebutkan minimal 2 perusahaan besar yang menggunakan ElasticSearch dalam proses bisnisnya! Di bagian mana ES tersebut digunakan?

Jawaban : Berikut perusahaan-perusahaan yang menggunakan ES yaitu Ruangguru, Tokopedia, Github, Udemy, dan Stack Overflow. Perusahaan-perusahaan tersebut rata-rata menggunakan ES pada bagian searching.

4. (20%) Bagaimana query dengan DSL untuk mencari user yang nama emailnya berawalan huruf a dan usianya di rentang 20 sampai 30 tahun?

Jawaban :

```
curl -X GET "localhost:9200/_search?pretty" -H 'Content-Type: application/json' -d'
{
  "query": {
    "query_string" : {
      "query" : "email: a* AND age: (>=20 AND <=30)"
    }
  }
}
```

5. (15%) Jelaskan apa itu Elastic Stack dan komponennya!

Jawaban : Elastic Stack merupakan susunan dari komponen berikut :

- Elasticsearch merupakan mesin pencarian dan analitik berbasis Apache Lucene yang scalable secara terstruktur ataupun tidak.
- Logstash merupakan apps untuk data collection pipeline yang mengumpulkan data dan memasukkannya ke dalam elasticsearch.
- Kibana merupakan dashboard untuk elasticsearch serta menyediakan eksplorasi secara visual dan analisis real-time terhadap data.

6. (30%) Diberikan sebuah kasus dalam dunia kesehatan untuk mendeteksi penyebaran penyakit tertentu dengan memanfaatkan Big Data.

(a) Data apa saja yang perlu dikumpulkan?

(b) Informasi/insight apa saja yang bisa didapatkan dari data tersebut?

(c) Ceritakan apa yang bisa dilakukan oleh perusahaan, pemerintah, atau masyarakat dari informasi/insight yang didapat tersebut!

Jawaban : Kasus terjadi dalam dunia kesehatan yaitu tersebarnya penyakit tertentu dan untuk mendeteksinya maka dimanfaatkan sebuah teknologi salah satunya Big Data. Sebelum mendeteksi penyebaran penyakitnya maka diperlukan data-data yang perlu dijadikan sebagai landasan untuk menentukan suatu keputusan, berikut data-data yang perlu dikumpulkan :

- Jumlah masyarakat yang terkena penyakit tersebut
- Kriteria masyarakat yang terkena penyakit tersebut
- Gejala pada masyarakat yang terkena penyakit tersebut

- Waktu penyakit tersebut menyebar di masyarakat
- Lokasi masyarakat yang terkena penyakit tersebut

Data di atas InsyaAllah cukup untuk mendeteksi penyebaran suatu penyakit dan dapat dijadikan informasi yang bermanfaat diantaranya :

- Mengetahui lokasi yang sudah terkena penyakit dengan jumlah masyarakat yang terinfeksi
- Mengetahui kapan penyakit tersebut menyebar dan dapat dilihat jumlah masyarakat yang terinfeksi selama beberapa waktu
- Dari data lokasi yang diketahui telah terinfeksi penyakit tersebut dapat dianalisis keadaan lokasi tersebut dari gejala dan kriteria masyarakatnya.
- Dari data waktu yang diketahui dapat mendeteksi tingkat penyebaran penyakit tersebut meningkat atau menurun.

Setelah diketahui informasi di atas, maka bisa dilakukan langkah-langkah penyebaran penyakit tersebut, seperti berikut :

- Menghindari lokasi yang ada penyebaran penyakit tersebut
- Menghindari masyarakat yang terinfeksi dengan memiliki gejala dan kriteria tertentu.
- Menghindari waktu pada saat penyakit tersebut meningkat pesat menyebar.