

Teknologi Penyimpanan Big Data

BIG DATA

ACID dan BASE

- ACID singkatan dari:
 - ✓ Atomicity
 - ✓ Consistency
 - ✓ Isolation
 - ✓ Durability
- BASE singkatan dari :
 - basically available
 - soft state
 - eventual consistency



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
tinggi

QS
STARS
RATING SYSTEM
2016

AKREDITASI
KUALITAS
NUSA

IAABE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Penyimpanan

- Teknologi penyimpanan (*storage*) terus berkembang dari waktu ke waktu, bergerak dari dalam server ke luar di jaringan.
- Kebutuhan untuk menyimpan Big Data telah mengubah pandangan relasional, database-sentris yang telah dianut oleh Enterprise ICT sejak akhir 1980-an.
- Intinya adalah bahwa teknologi relasional sama sekali tidak dapat diskalakan untuk mendukung volume Big Data.
- Bisnis dapat menemukan nilai murni dalam memproses data semi-terstruktur dan tidak terstruktur, yang umumnya tidak sesuai dengan pendekatan relasional.
- Big Data telah mendorong batas penyimpanan ke tampilan terpadu dari memori yang tersedia dan penyimpanan disk dari sebuah cluster.
- Jika diperlukan lebih banyak penyimpanan, skalabilitas horizontal memungkinkan perluasan cluster melalui penambahan lebih banyak node.



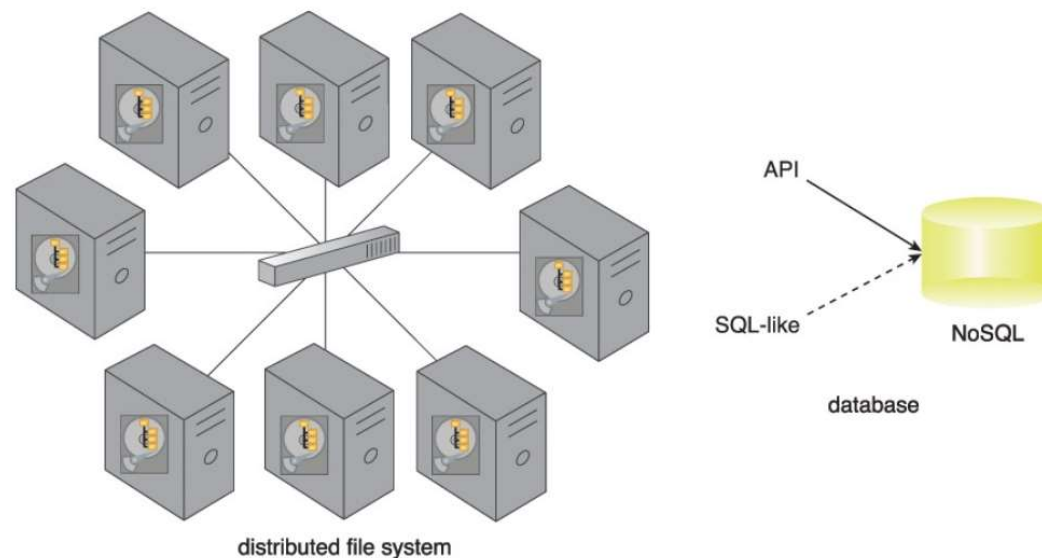
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

On-Disk Storage Devices

- Penyimpanan pada disk umumnya menggunakan *hard disk drive* berbiaya rendah untuk penyimpanan jangka panjang.
- On-disk storage dapat diimplementasikan melalui sistem file terdistribusi atau database.



Sistem File Terdistribusi

- Sistem file terdistribusi, seperti sistem file lainnya, tidak tahu (agnostic) data yang disimpan dan oleh karena itu mendukung penyimpanan data tanpa skema.
- Secara umum, perangkat penyimpanan sistem file terdistribusi menyediakan redundansi out-of-box dan ketersediaan tinggi dengan menyalin data ke beberapa lokasi melalui replikasi.
- Perangkat penyimpanan yang diimplementasikan dengan sistem file terdistribusi menyediakan penyimpanan data akses cepat dan sederhana yang mampu menyimpan kumpulan data besar yang bersifat non-relasional, seperti data semi terstruktur dan tidak terstruktur.
- Meskipun berdasarkan mekanisme penguncian file langsung untuk kontrol konkurensi, sistem file terdistribusi menyediakan kemampuan baca/tulis cepat, yang mendukung karakteristik kecepatan Big Data.
- Sistem file terdistribusi tidak ideal untuk kumpulan data yang terdiri dari sejumlah besar file kecil karena menciptakan aktivitas pencarian disk yang berlebihan, memperlambat akses data secara keseluruhan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Sistem File Terdistribusi

- Ada juga lebih banyak overhead yang terlibat dalam pemrosesan beberapa file yang lebih kecil, karena proses khusus umumnya dihasilkan oleh mesin pemrosesan saat runtime untuk memproses setiap file sebelum hasilnya disinkronkan dari seluruh cluster.
- Karena keterbatasan ini, sistem file terdistribusi bekerja paling baik dengan file yang lebih sedikit tetapi lebih besar diakses secara berurutan.
- Beberapa file yang lebih kecil umumnya digabungkan menjadi satu file untuk memungkinkan penyimpanan dan pemrosesan yang optimal.
- Hal ini memungkinkan sistem file terdistribusi untuk meningkatkan kinerja ketika data harus diakses dalam mode streaming tanpa membaca dan menulis secara acak.



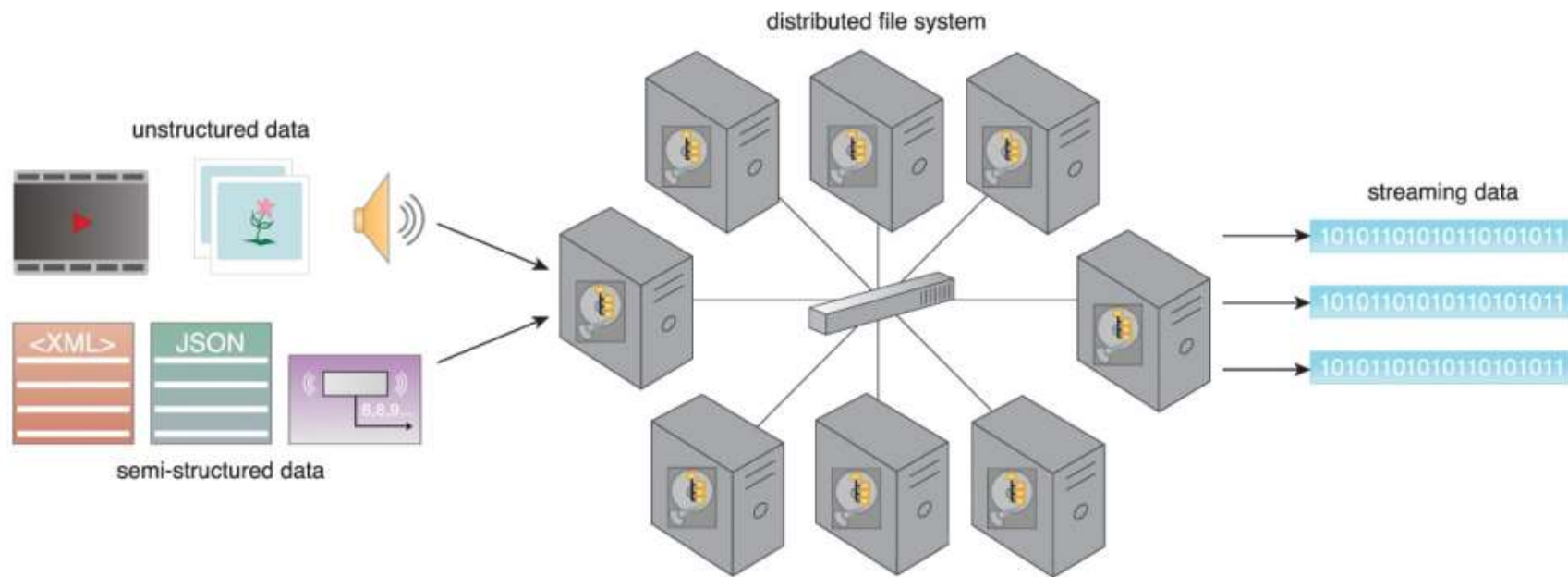
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Sistem File Terdistribusi

Sistem file terdistribusi yang mengakses data dalam mode streaming tanpa pembacaan dan penulisan acak.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN PT

A
tinggi

QS STARS
RATING SYSTEM
2020

ICAC
INDONESIA

IAEE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Sistem File Terdistribusi

- Perangkat penyimpanan sistem file terdistribusi cocok ketika kumpulan data besar dari data mentah akan disimpan atau ketika pengarsipan kumpulan data diperlukan.
- Sistem file terdistribusi menyediakan opsi penyimpanan yang murah untuk menyimpan data dalam jumlah besar dalam jangka waktu yang lama yang harus tetap online.
- Ini karena lebih banyak disk dapat dengan mudah ditambahkan ke cluster tanpa perlu memindahkan data ke penyimpanan data offline, seperti kaset.
- Perlu dicatat bahwa sistem file terdistribusi tidak menyediakan kemampuan untuk mencari isi file sebagai kemampuan standar out-of-the-box.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

RDBMS Databases

- Relational database management systems (RDBMSs) baik untuk menangani beban kerja transaksional yang melibatkan sejumlah kecil data dengan properti baca/tulis acak.
- RDBMS sesuai dengan ACID, umumnya dibatasi untuk satu node.
- RDBMS tidak menyediakan redundansi yang siap pakai dan toleransi kesalahan.
- Untuk menangani sejumlah besar data yang tiba dengan cepat, database relasional umumnya perlu diskalakan.
- RDBMS menggunakan penskalaan vertikal, bukan penskalaan horizontal, yang merupakan strategi penskalaan yang lebih mahal dan mengganggu.
- Hal ini membuat RDBMS kurang ideal untuk penyimpanan data jangka panjang yang terakumulasi dari waktu ke waktu.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

RDBMS Databases

- Perhatikan bahwa beberapa database relasional, misalnya IBM DB2 pureScale, Sybase ASE Cluster Edition, Oracle Real Application Clusters (RAC) dan Microsoft Parallel Data Warehouse (PDW), mampu dijalankan pada cluster.
- Namun, cluster database ini masih menggunakan penyimpanan bersama yang dapat bertindak sebagai titik kegagalan tunggal.
- Sebuah cluster Basis data relasional menggunakan arsitektur penyimpanan bersama, yang memungkinkan satu titik kegagalan yang memengaruhi ketersediaan basis data.

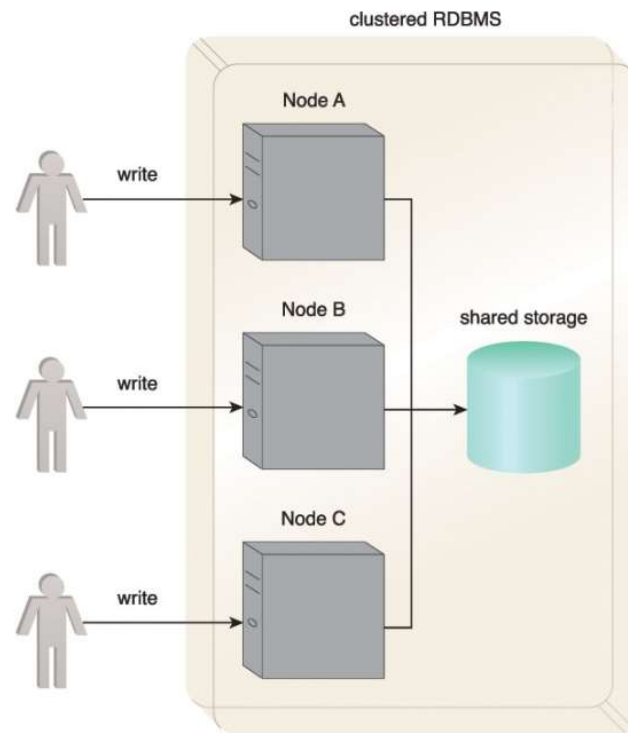


UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

RDBMS Databases



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

RDBMS Databases

- Basis data relasional perlu dilakukan sharding secara manual, sebagian besar menggunakan logika aplikasi.
- Ini berarti bahwa logika aplikasi perlu mengetahui shard mana yang harus dikueri untuk mendapatkan data yang diperlukan.
- Ini semakin memperumit pemrosesan data ketika data dari beberapa shard diperlukan.

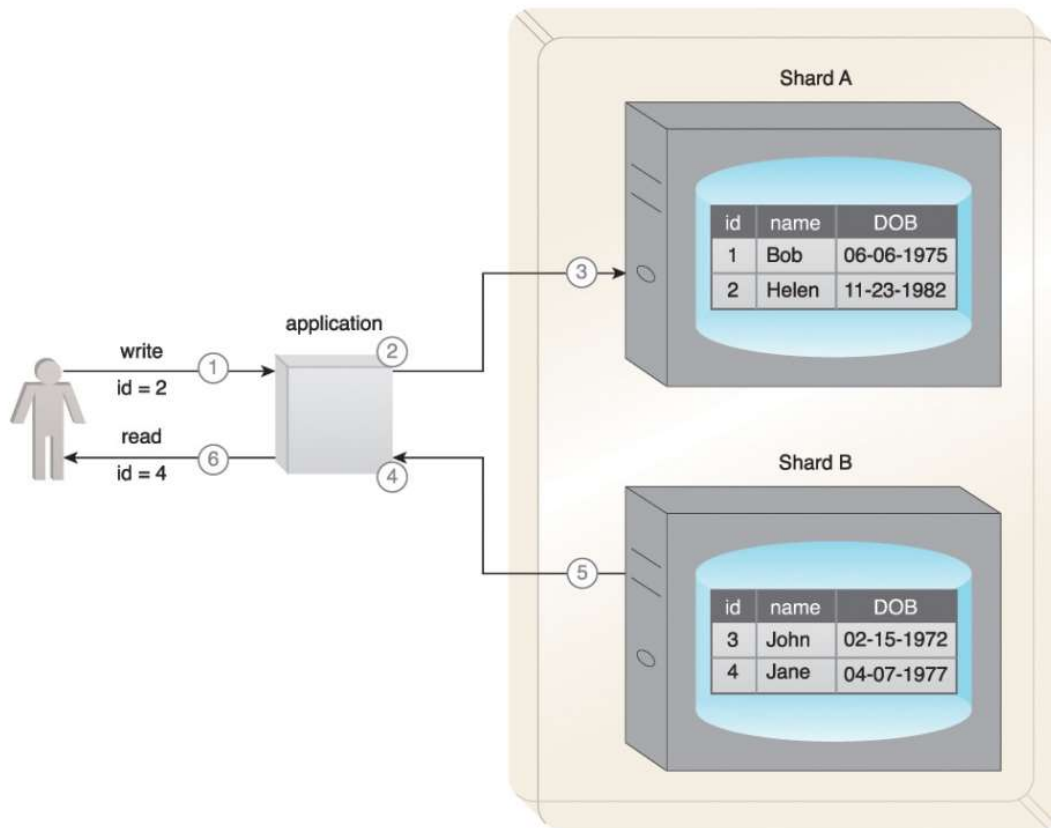


UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Basis data relasional dipecah secara manual menggunakan logika aplikasi.



1. Seorang pengguna menulis sebuah *record* (id = 2).
2. Logika aplikasi menentukan shard mana yang harus ditulis.
3. *Record* dikirim ke pecahan yang ditentukan oleh logika aplikasi.
4. Pengguna membaca catatan (id = 4), dan logika aplikasi menentukan pecahan mana yang berisi data.
5. Data dibaca dan dikembalikan ke aplikasi.
6. Aplikasi kemudian mengembalikan *record* ke pengguna.

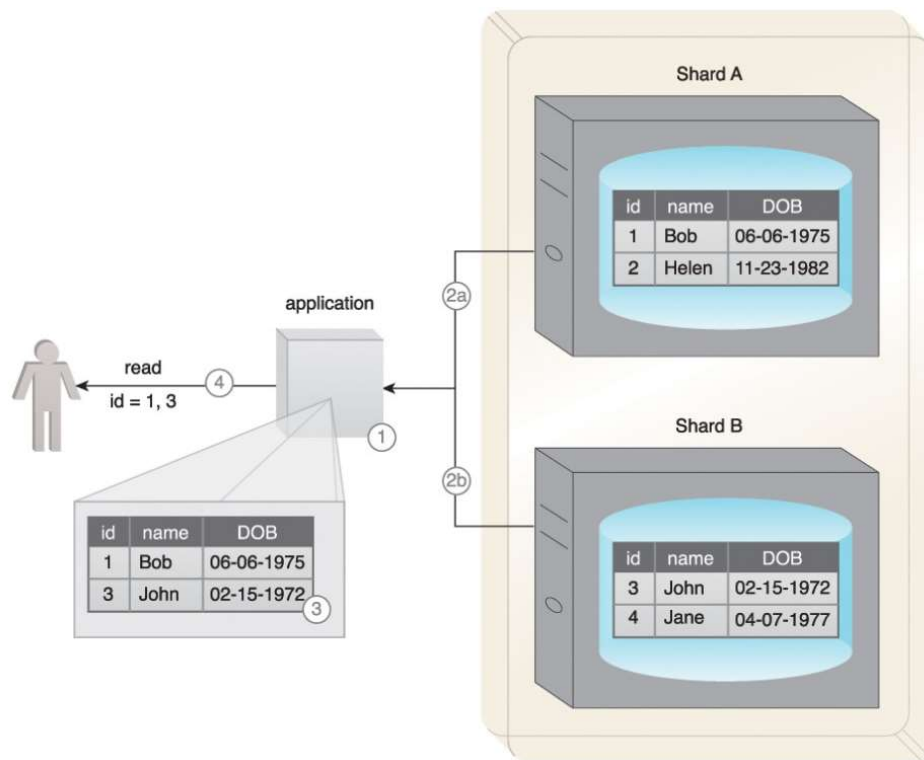


UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Contoh penggunaan logika aplikasi untuk menggabungkan data yang diambil dari beberapa pecahan.



1. Seorang pengguna meminta beberapa *record* (id = 1, 3) dan logika aplikasi digunakan untuk menentukan pecahan mana yang perlu dibaca.
2. Ditentukan oleh logika aplikasi bahwa Shard A dan Shard B perlu dibaca.
3. Data dibaca dan digabungkan oleh aplikasi.
4. Akhirnya, data dikembalikan ke pengguna.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

RDBMS Databases

- Basis data relasional umumnya membutuhkan data untuk mematuhi skema.
- Akibatnya, penyimpanan data semi-terstruktur dan tidak terstruktur yang skemanya non-relasional tidak didukung secara langsung.
- Selanjutnya, dengan skema database relasional kesesuaian divalidasi pada saat memasukkan atau memperbarui data dengan memeriksa data terhadap batasan skema.
- Ini memperkenalkan overhead yang menciptakan latensi.
- Latensi ini membuat database relasional menjadi pilihan yang kurang ideal untuk menyimpan data berkecepatan tinggi yang membutuhkan perangkat penyimpanan database yang tersedia dengan kemampuan menulis data yang cepat.
- Sebagai akibat dari kekurangannya, RDBMS tradisional umumnya tidak berguna sebagai perangkat penyimpanan utama dalam lingkungan solusi Big Data.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

NoSQL Databases

- Not Only SQL (NoSQL) mengacu pada teknologi yang digunakan untuk mengembangkan database nonrelasional generasi berikutnya yang sangat skalabel dan toleran terhadap kesalahan.
- Simbol untuk merepresentasikan basis data NoSQL.



NoSQL Database



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Karakteristik NoSQL

- Schema-less data model – Data bisa ada dalam bentuk mentahnya.
- Scale-out rather than scale up – Lebih banyak node dapat ditambahkan untuk mendapatkan penyimpanan tambahan dengan database NoSQL, berbeda dengan harus mengganti node yang ada dengan yang lebih baik, kinerja/kapasitas lebih tinggi.
- Highly available – Dibangun di atas teknologi berbasis cluster yang memberikan toleransi kesalahan out of the box.
- Lower operational costs – Banyak database NoSQL dibangun di atas platform Open Source tanpa biaya lisensi. Mereka sering dapat digunakan pada perangkat keras komoditas.
- Eventual consistency – Data dibaca di beberapa node tetapi mungkin tidak konsisten segera setelah penulisan. Namun, semua node pada akhirnya akan berada dalam kondisi yang konsisten.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Karakteristik NoSQL

- BASE, not ACID – Kepatuhan BASE memerlukan database untuk menjaga ketersediaan tinggi jika terjadi kegagalan jaringan/node, sementara tidak mengharuskan database berada dalam keadaan yang konsisten setiap kali pembaruan terjadi. Basis data dapat dalam keadaan lunak/tidak konsisten hingga akhirnya mencapai konsistensi. Akibatnya, dengan mempertimbangkan teorema CAP, perangkat penyimpanan NoSQL umumnya AP atau CP.
- API driven data access – Akses data umumnya didukung melalui kueri berbasis API, termasuk RESTful API, sedangkan beberapa implementasi juga dapat menyediakan kemampuan kueri seperti SQL
- Auto sharding and replication – Untuk mendukung penskalaan horizontal dan menyediakan ketersediaan tinggi, perangkat penyimpanan NoSQL secara otomatis menggunakan teknik sharding dan replikasi di mana kumpulan data dipartisi secara horizontal dan kemudian disalin ke beberapa node.
- Integrated caching – Ini menghilangkan kebutuhan akan lapisan caching terdistribusi pihak ketiga, seperti Memcached.



Karakteristik NoSQL

- Distributed query support – Perangkat penyimpanan NoSQL mempertahankan perilaku kueri yang konsisten di beberapa shard.
- Polyglot persistence – Penggunaan penyimpanan NoSQL tidak mengharuskan penghentian RDBMS tradisional. Faktanya, keduanya dapat digunakan pada saat yang sama, yang merupakan pendekatan untuk mempertahankan data menggunakan berbagai jenis teknologi penyimpanan dalam arsitektur solusi yang sama. Ini bagus untuk mengembangkan sistem yang membutuhkan data terstruktur dan semi/tidak terstruktur.
- Aggregate-focused – Tidak seperti database relasional yang paling efektif dengan data yang sepenuhnya dinormalisasi, perangkat penyimpanan NoSQL menyimpan data agregat yang tidak dinormalisasi (entitas yang berisi data gabungan, sering kali bersarang, untuk suatu objek) sehingga menghilangkan kebutuhan untuk bergabung dan pemetaan ekstensif antar aplikasi objek dan data yang disimpan dalam database. Satu pengecualian, perangkat penyimpanan basis data grafik (diperkenalkan segera) tidak berfokus pada agregat.
- Catatan “Tidak semua perangkat penyimpanan NoSQL menunjukkan semua fitur ini.”



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

NoSQL

- Rationale

- Munculnya perangkat penyimpanan NoSQL terutama dapat dikaitkan dengan karakteristik volume, kecepatan, dan variasi dari kumpulan data Big Data.

- Volume

- Persyaratan penyimpanan volume data yang terus meningkat memerintahkan penggunaan database yang sangat skalabel sambil menekan biaya agar bisnis tetap kompetitif.
- Perangkat penyimpanan NoSQL memenuhi persyaratan ini dengan menyediakan kemampuan scale out saat menggunakan server komoditas yang murah.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

NoSQL

- Velocity

- Aliran data yang cepat membutuhkan database dengan kemampuan menulis data akses cepat.
- Perangkat penyimpanan NoSQL memungkinkan penulisan cepat dengan menggunakan prinsip schema-on-read daripada schema-on-write.
- Menjadi *highly available*, perangkat penyimpanan NoSQL dapat memastikan bahwa latensi tulis tidak terjadi karena kegagalan node atau jaringan.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
Linggi

QS STARS
RATING SYSTEM
5 Star

AMBA
ACCREDITED

IAEE

CPA
AUSTRALIA

ICAEW

CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

NoSQL

- Variety

- Perangkat penyimpanan perlu menangani berbagai format data termasuk dokumen, email, gambar dan video, serta data yang tidak lengkap.
- Perangkat penyimpanan NoSQL dapat menyimpan berbagai bentuk format data semi-terstruktur dan tidak terstruktur.
- Pada saat yang sama, perangkat penyimpanan NoSQL dapat menyimpan data tanpa skema dan data yang tidak lengkap dengan kemampuan tambahan untuk membuat perubahan skema seiring dengan berkembangnya model data dari kumpulan data.
- Dengan kata lain, database NoSQL mendukung evolusi skema.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

NoSQL

- Types

- Perangkat penyimpanan NoSQL terutama dapat dibagi menjadi empat jenis berdasarkan caranya menyimpan data.
 - key-value
 - document
 - column-family
 - graph



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
Lingguh

QS
STARS
RATING SYSTEM
2016

AKREDITASI
KUALITAS
NUSA

IAABE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

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

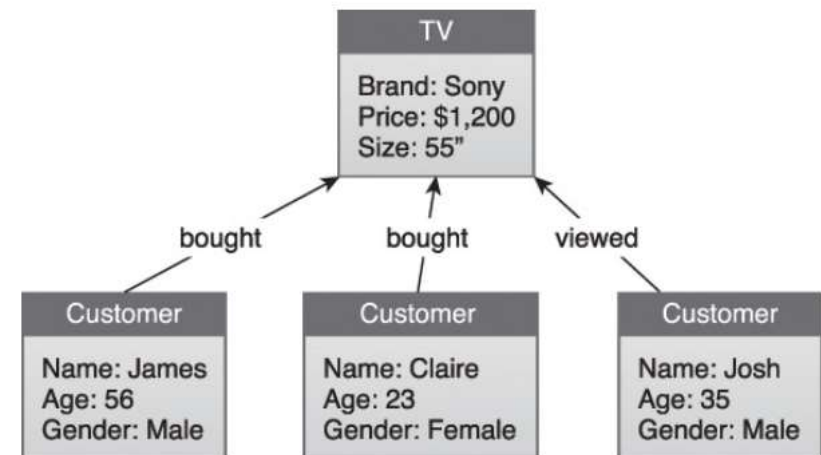
Contoh key-value NoSQL storage



Contoh document NoSQL storage

studentId	personal details	address	modules history
821	FirstName: Cristie LastName: Augustin DoB: 03-15-1992 Gender: Female Ethnicity: French	Street: 123 New Ave City: Portland State: Oregon ZipCode: 12345 Country: USA	Taken: 5 Passed: 4 Failed: 1
742	FirstName: Carlos LastName: Rodriguez MiddleName: Jose Gender: Male	Street: 456 Old Ave City: Los Angeles Country: USA	Taken: 7 Passed: 5 Failed: 2

Contoh column-family NoSQL storage



Contoh graph NoSQL storage



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Key-Value

- Perangkat penyimpanan key-value menyimpan data sebagai pasangan nilai kunci dan bertindak seperti tabel hash.
- Tabel adalah daftar nilai di mana setiap nilai diidentifikasi dengan kunci.
- Nilainya “kabur” ke database dan biasanya disimpan sebagai BLOB.
- Nilai yang disimpan dapat berupa agregat apa pun, mulai dari data sensor hingga video.
- Pencarian nilai hanya dapat dilakukan melalui kunci karena database tidak menyadari detail agregat yang disimpan.
- Pembaruan Sebagian (Partial updates) tidak dimungkinkan.
- Pembaruan (update) adalah operasi delete atau insert.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Key-Value

- Perangkat penyimpanan key-value umumnya tidak memelihara indeks apa pun, oleh karena itu penulisan cukup cepat.
- Berdasarkan model penyimpanan sederhana, perangkat penyimpanan key-value sangat skalabel.
- Karena kunci (key) adalah satu-satunya cara untuk mengambil data, kunci biasanya ditambahkan dengan jenis nilai yang disimpan untuk pengambilan yang mudah.
- Contohnya adalah 123_sensor1.
- Untuk menyediakan beberapa struktur pada data yang disimpan, sebagian besar perangkat penyimpanan key-value menyediakan *collections* atau *buckets* (seperti tabel) di mana pasangan key-value dapat diatur.
- Contoh perangkat penyimpanan key-value yaitu Riak, Redis, dan Amazon Dynamodb.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Key-Value

- Sebuah *collections* dapat menampung banyak format data.
- Beberapa implementasi mendukung nilai kompresi untuk mengurangi jejak penyimpanan.
- Namun, ini memperkenalkan latensi pada waktu baca, karena data perlu didekompresi terlebih dahulu sebelum dikembalikan.
- Contoh data yang diatur ke dalam pasangan key-value.

key	value
631	John Smith, 10.0.30.25, Good customer service
365	10101101010110101011010110101011010101110101110
198	<CustomerId>32195</CustomerId><Total>43.25</Total>

← text

← image

← XML

Key-Value

- Perangkat penyimpanan key-value baik digunakan jika:
 - penyimpanan data tidak terstruktur diperlukan
 - membaca/menulis kinerja tinggi diperlukan
 - nilainya dapat diidentifikasi sepenuhnya melalui kunci saja
 - nilai adalah entitas mandiri yang tidak bergantung pada nilai lain
 - nilai memiliki struktur yang relatif sederhana atau kueri biner
 - polanya sederhana, hanya melibatkan operasi insert, select, dan delete
 - nilai yang disimpan dimanipulasi pada lapisan aplikasi



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
tinggi

QS
STARS
RATING SYSTEM
5 star

ISAS
ISAS

IAEE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Key-Value

- Perangkat penyimpanan nilai kunci tidak tepat digunakan jika:
 - aplikasi memerlukan pencarian atau penyaringan data menggunakan atribut dari nilai yang disimpan
 - ada hubungan antara entri nilai kunci yang berbeda
 - sekelompok nilai kunci perlu diperbarui dalam satu transaksi
 - beberapa kunci memerlukan manipulasi dalam satu operasi
 - konsistensi skema di seluruh nilai yang berbeda diperlukan
 - pembaruan ke atribut individual dari nilai diperlukan



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Dokumen

- Perangkat penyimpanan dokumen juga menyimpan data sebagai pasangan key-value.
- Namun, tidak seperti perangkat penyimpanan key-value, nilai yang disimpan adalah dokumen yang dapat ditanyakan oleh database.
- Dokumen-dokumen ini dapat memiliki struktur bersarang yang kompleks, seperti faktur.
- Dokumen dapat dikodekan menggunakan skema pengkodean berbasis teks, seperti XML atau JSON, atau menggunakan skema pengkodean biner, seperti BSON (Binary JSON).
- Contoh perangkat penyimpanan dokumen yaitu MongoDB, CouchDB, dan Terrastore.

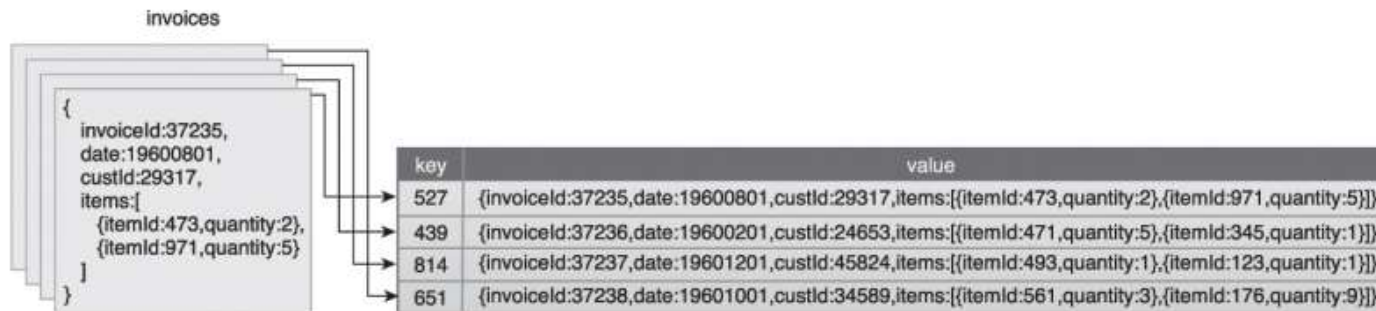


UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Dokumen



Penggambaran data JSON yang disimpan dalam perangkat penyimpanan dokumen.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Dokumen

- Seperti perangkat penyimpanan key-value, sebagian besar perangkat penyimpanan dokumen menyediakan *collections* atau *bucket* (seperti tabel) di mana pasangan key-value dapat diatur.
- Perbedaan utama antara perangkat penyimpanan dokumen dan perangkat penyimpanan key-value adalah sebagai berikut:
 - ✓ perangkat penyimpanan dokumen *value-aware*.
 - ✓ nilai yang disimpan menggambarkan dirinya sendiri; skema dapat disimpulkan dari struktur nilai atau referensi ke skema untuk dokumen yang termasuk dalam nilai.
 - ✓ operasi *select* dapat mereferensikan bidang di dalam nilai agregat.
 - ✓ operasi *select* dapat mengambil bagian dari nilai agregat.
 - ✓ *partial update* didukung; oleh karena itu sebagian dari agregat dapat diperbarui.
 - ✓ indeks yang mempercepat pencarian umumnya didukung.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Dokumen

- Setiap dokumen dapat memiliki skema yang berbeda; oleh karena itu, dimungkinkan untuk menyimpan berbagai jenis dokumen dalam *collection* atau *bucket* yang sama.
- *Fields* tambahan dapat ditambahkan ke dokumen setelah penyisipan awal, sehingga memberikan dukungan skema yang fleksibel.
- Perlu dicatat bahwa perangkat penyimpanan dokumen tidak terbatas pada penyimpanan data yang terjadi dalam bentuk dokumen aktual, seperti file XML, tetapi juga dapat digunakan untuk menyimpan agregat apa pun yang terdiri dari kumpulan *fields* yang memiliki skema *flat* atau *nested schema* (skema bersarang).



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Dokumen

- Perangkat penyimpanan dokumen baik digunakan jika:
 - menyimpan data berorientasi dokumen semi-terstruktur yang terdiri dari skema datar (flat-schema) atau bertingkat (nested-schema)
 - evolusi skema merupakan syarat karena struktur dokumen tidak diketahui atau kemungkinan akan berubah
 - aplikasi memerlukan pembaruan sebagian dari agregat yang disimpan sebagai dokumen
 - pencarian perlu dilakukan di berbagai bidang dokumen
 - menyimpan objek domain, seperti pelanggan, dalam bentuk objek berseri
 - pola kueri melibatkan operasi insert, select, update, dan delete



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Dokumen

- Perangkat penyimpanan dokumen tidak sesuai digunakan jika:
 - beberapa dokumen perlu diperbarui sebagai bagian dari satu transaksi
 - melakukan operasi yang memerlukan penggabungan antara beberapa dokumen atau menyimpan data yang dinormalisasi
 - penegakan skema untuk mencapai desain kueri yang konsisten diperlukan karena struktur dokumen dapat berubah antara menjalankan kueri yang berurutan, yang akan memerlukan restrukturisasi kueri
 - nilai yang disimpan tidak menggambarkan dirinya sendiri dan tidak memiliki referensi ke skema
 - data biner perlu disimpan



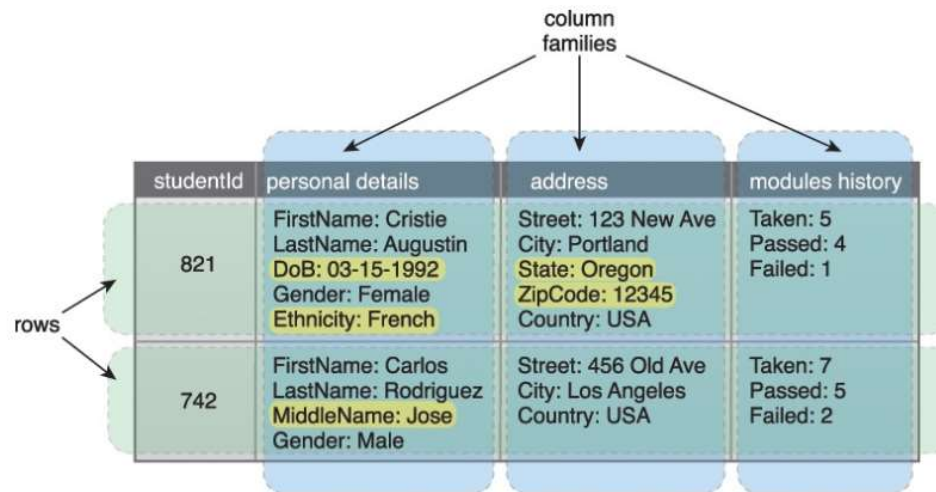
UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Column-Family

- Perangkat penyimpanan column-family menyimpan data seperti RDBMS tradisional tetapi mengelompokkan kolom terkait bersama-sama dalam satu baris, menghasilkan keluarga kolom.
- Setiap kolom dapat berupa kumpulan kolom terkait itu sendiri, yang disebut sebagai super-kolom.
- Kolom yang disorot menggambarkan fitur skema fleksibel yang didukung oleh database column-family, di mana setiap baris dapat memiliki kumpulan kolom yang berbeda.



Column-Family

- Setiap super-kolom dapat berisi sejumlah kolom terkait yang sewenang-wenang yang umumnya diambil atau diperbarui sebagai satu unit.
- Setiap baris terdiri dari beberapa keluarga kolom dan dapat memiliki kumpulan kolom yang berbeda, sehingga menunjukkan dukungan skema yang fleksibel.
- Setiap baris diidentifikasi oleh kunci baris.
- Perangkat penyimpanan column-family menyediakan akses data cepat dengan kemampuan baca/tulis acak.
- Mereka menyimpan kelompok kolom yang berbeda dalam file fisik terpisah, yang meningkatkan respon kueri karena hanya kelompok kolom yang diperlukan yang dicari.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Column-Family

- Beberapa perangkat penyimpanan column-family menyediakan dukungan untuk mengompresi column-family secara selektif.
- Membiarkan kelompok kolom yang dapat ditelusuri tidak terkompresi dapat membuat kueri lebih cepat karena kolom target tidak perlu didekompresi untuk pencarian.
- Sebagian besar implementasi mendukung pembuatan versi data sementara beberapa mendukung menentukan waktu kedaluwarsa untuk data kolom.
- Ketika waktu kedaluwarsa telah berlalu, data secara otomatis dihapus.
- Contoh perangkat penyimpanan column-family yaitu Cassandra, HBase, dan Amazon SimpleDB.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Column-Family

- Perangkat penyimpanan column-family sesuai digunakan jika:
 - kemampuan baca/tulis acak real-time diperlukan dan data yang disimpan memiliki beberapa struktur yang ditentukan
 - data mewakili struktur tabular, setiap baris terdiri dari sejumlah besar kolom dan kelompok bersarang dari data yang saling terkait ada
 - dukungan untuk evolusi skema diperlukan karena kelompok kolom dapat ditambahkan atau dihapus tanpa waktu henti sistem
 - *fields* tertentu sebagian besar diakses bersama-sama, dan pencarian perlu dilakukan menggunakan nilai bidang
 - penggunaan penyimpanan yang efisien diperlukan ketika data terdiri dari baris yang isinya jarang karena database column-family hanya mengalokasikan ruang penyimpanan jika ada kolom untuk satu baris. Jika tidak ada kolom, tidak ada ruang yang dialokasikan.
 - pola kueri melibatkan operasi insert, select, update, dan delete



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Column-Family

- Perangkat penyimpanan column-family tidak sesuai digunakan jika:
 - akses data relasional diperlukan; misalnya joins
 - dukungan transaksi ACID diperlukan
 - data biner perlu disimpan
 - kueri yang sesuai dengan SQL harus dijalankan
 - pola kueri cenderung sering berubah karena hal itu dapat memulai restrukturisasi terkait tentang bagaimana column-family diatur



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Graph

- Perangkat penyimpanan grafik digunakan untuk mempertahankan entitas yang saling terhubung.
- Tidak seperti perangkat penyimpanan NoSQL lainnya, di mana penekanannya adalah pada struktur entitas, perangkat penyimpanan grafik menekankan pada penyimpanan hubungan antar entitas.
- Entitas disimpan sebagai node (jangan tertukar dengan node cluster) dan juga disebut vertex, sedangkan hubungan antar entitas disimpan sebagai edge.
- Dalam bahasa RDBMS, setiap node dapat dianggap sebagai satu baris sementara edge menunjukkan join.
- Node dapat memiliki lebih dari satu jenis tautan melalui banyak edges.
- Setiap node dapat memiliki data atribut sebagai pasangan key-value, seperti node pelanggan dengan atribut ID, nama, dan usia.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Graph

- Setiap edge dapat memiliki data atributnya sendiri sebagai pasangan key-value, yang dapat digunakan untuk memfilter hasil kueri lebih lanjut.
- Memiliki banyak edge mirip dengan mendefinisikan beberapa foreign key dalam RDBMS; namun, tidak setiap node diharuskan memiliki edge yang sama.
- Query umumnya melibatkan pencarian node yang saling berhubungan berdasarkan atribut node dan/atau atribut edge, yang biasa disebut sebagai node traversal.
- Edge bisa searah atau dua arah, mengatur arah traversal node.
- Umumnya, perangkat penyimpanan grafik memberikan konsistensi melalui kepatuhan ACID.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Graph

- Tingkat kegunaan perangkat penyimpanan grafik tergantung pada jumlah dan jenis edge yang didefinisikan di antara node.
- Semakin besar jumlahnya dan semakin beragam edge, semakin beragam jenis kueri yang dapat ditanganinya.
- Akibatnya, penting untuk secara komprehensif menangkap jenis hubungan yang ada antara node.
- Ini tidak hanya berlaku untuk skenario penggunaan yang ada, tetapi juga untuk analisis eksplorasi data.
- Perangkat penyimpanan grafik umumnya memungkinkan penambahan jenis node baru tanpa membuat perubahan pada database.
- Ini juga memungkinkan pendefinisian link tambahan antara node sebagai tipe baru dari hubungan atau node yang muncul di database.
- Contoh penyimpanan graph yaitu Neo4J, Infinite Graph, dan OrientDB.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

Graph

- Perangkat penyimpanan grafik sesuai digunakan jika:
 - entitas yang saling berhubungan perlu disimpan
 - menanyakan entitas berdasarkan jenis hubungan satu sama lain daripada atribut entitas
 - menemukan kelompok entitas yang saling berhubungan
 - menemukan jarak antara entitas dalam hal jarak traversal node
 - menambang data dengan tujuan untuk menemukan pola



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
Linggi

QS
STARS
RATING SYSTEM
2020

CSA
KUALITAS
SISTEM

IAEE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

Graph

- Perangkat penyimpanan grafik tidak sesuai digunakan jika:
 - update diperlukan untuk sejumlah besar atribut node atau atribut edge, karena ini melibatkan pencarian node atau edge, yang merupakan operasi yang mahal dibandingkan dengan melakukan traversal node
 - entitas memiliki banyak atribut atau data bersarang (nested data)—yang terbaik adalah menyimpan entitas ringan di perangkat penyimpanan grafik sambil menyimpan data atribut lainnya di perangkat penyimpanan NoSQL non-grafik yang terpisah
 - penyimpanan biner diperlukan
 - query berdasarkan pemilihan atribut node/edge mendominasi query traversal node



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

NewSQL Databases

- Perangkat penyimpanan NoSQL sangat skalabel, tersedia, toleran terhadap kesalahan, dan cepat untuk operasi baca/tulis.
- Namun, mereka tidak memberikan dukungan transaksi dan konsistensi yang sama seperti yang ditunjukkan oleh ACID yang sesuai RDBMS.
- Mengikuti model BASE, perangkat penyimpanan NoSQL memberikan konsistensi akhir daripada konsistensi langsung.
- Oleh karena itu mereka akan berada dalam keadaan lunak (soft state) saat mencapai keadaan konsistensi akhirnya.
- Akibatnya, mereka tidak sesuai untuk digunakan saat menerapkan sistem transaksional skala besar.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

NewSQL Databases

- Perangkat penyimpanan NewSQL menggabungkan properti ACID dari RDBMS dengan skalabilitas dan toleransi kesalahan yang ditawarkan oleh perangkat penyimpanan NoSQL.
- Database NewSQL umumnya mendukung sintaks yang sesuai dengan SQL untuk definisi data dan operasi manipulasi data, dan mereka sering menggunakan model data relasional logis untuk penyimpanan data.
- Database NewSQL dapat digunakan untuk mengembangkan sistem OLTP dengan volume transaksi yang sangat tinggi, misalnya sistem perbankan.
- Mereka juga dapat digunakan untuk analitik real-time, misalnya analitik operasional, karena beberapa implementasi memanfaatkan penyimpanan dalam memori.
- Dibandingkan dengan perangkat penyimpanan NoSQL, perangkat penyimpanan NewSQL menyediakan transisi yang lebih mudah dari RDBMS tradisional ke database yang sangat skalabel karena dukungannya untuk SQL.
- Contoh database NewSQL termasuk VoltDB, NuoDB dan InnoDB.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Storage Devices

- Bagian sebelumnya memperkenalkan perangkat penyimpanan on-disk dan berbagai jenisnya sebagai sarana dasar penyimpanan data.
- Bagian ini dibangun di atas pengetahuan dengan menghadirkan penyimpanan dalam memori sebagai sarana untuk menyediakan opsi untuk penyimpanan data tingkat lanjut yang berkinerja tinggi.
- Perangkat penyimpanan dalam memori umumnya menggunakan RAM, memori utama komputer, sebagai media penyimpanannya untuk menyediakan akses data yang cepat.
- Peningkatan kapasitas dan penurunan biaya RAM, ditambah dengan peningkatan kecepatan baca/tulis hard drive solid state, telah memungkinkan untuk mengembangkan solusi penyimpanan data dalam memori.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Storage Devices

- Penyimpanan data dalam memori menghilangkan latensi I/O disk dan waktu transfer data antara memori utama dan hard drive.
- Pengurangan keseluruhan dalam latensi baca/tulis data ini membuat pemrosesan data menjadi lebih cepat.
- Kapasitas perangkat penyimpanan dalam memori dapat ditingkatkan secara besar-besaran dengan menskalakan secara horizontal cluster yang menampung perangkat penyimpanan dalam memori.
- Memori berbasis cluster memungkinkan penyimpanan data dalam jumlah besar, termasuk kumpulan data Big Data, yang dapat diakses jauh lebih cepat jika dibandingkan dengan perangkat penyimpanan di disk.
- Ini secara signifikan mengurangi waktu eksekusi keseluruhan analitik Big Data, sehingga memungkinkan analitik Big Data secara realtime.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Storage Devices

- Gambar 1 menunjukkan bahwa pembacaan berurutan sebesar 1 MB data dari perangkat penyimpanan dalam memori membutuhkan waktu sekitar 0,25 ms.
- Gambar 2 menunjukkan bahwa membaca jumlah data yang sama dari perangkat penyimpanan pada disk membutuhkan waktu sekitar 20 ms.



Gambar 1



Gambar 2

UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
Simpul

QS
STARS
RATING SYSTEM

ISO
9001

ISO
14001

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

In-Memory Storage Devices

- Ini menunjukkan bahwa membaca data dari penyimpanan dalam memori kira-kira 80 kali lebih cepat daripada penyimpanan di disk.
- Diasumsikan bahwa waktu transfer data jaringan sama di kedua skenario dan karena itu telah dikeluarkan dari waktu baca.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Storage Devices

- Perangkat penyimpanan dalam memori memungkinkan analitik dalam memori, yang mengacu pada analisis data dalam memori, seperti menghasilkan statistik dengan mengeksekusi kueri pada data yang disimpan dalam memori, bukan di disk.
- Analitik dalam memori memungkinkan analitik operasional dan BI operasional melalui eksekusi kueri dan algoritme yang cepat.
- Terutama, penyimpanan dalam memori memungkinkan memahami masuknya data dengan cepat di lingkungan Big Data (karakteristik kecepatan) dengan menyediakan media penyimpanan yang memfasilitasi pembuatan informasi real-time.
- Ini mendukung pengambilan keputusan bisnis yang cepat untuk mengurangi ancaman atau memanfaatkan peluang.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Storage Devices

- Perangkat penyimpanan dalam memori Big Data diimplementasikan pada sebuah cluster, memberikan ketersediaan dan redundansi yang tinggi.
- Oleh karena itu, skalabilitas horizontal dapat dicapai hanya dengan menambahkan lebih banyak node atau memori.
- Jika dibandingkan dengan perangkat penyimpanan pada disk, perangkat penyimpanan dalam memori lebih mahal karena biaya memori yang lebih tinggi dibandingkan dengan perangkat penyimpanan berbasis disk.
- Meskipun mesin 64-bit dapat menggunakan memori 16 exabyte, karena keterbatasan fisik mesin, seperti jumlah ruang memori, memori yang terpasang jauh lebih sedikit.
- Untuk scaling out, tidak hanya penambahan memori, tetapi juga penambahan node yang diperlukan setelah batas memori per node tercapai. Hal ini meningkatkan biaya penyimpanan data.



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
tinggi

QS
STARS
RATING SYSTEM

ISAS
IABEE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

In-Memory Storage Devices

- Selain mahal, perangkat penyimpanan dalam memori tidak memberikan tingkat dukungan yang sama untuk penyimpanan data yang tahan lama.
- Faktor harga selanjutnya mempengaruhi kapasitas yang dapat dicapai dari perangkat dalam memori bila dibandingkan dengan perangkat penyimpanan pada disk.
- Akibatnya, hanya data lebih baru dan terbaru yang memiliki nilai paling tinggi yang disimpan dalam memori, sedangkan data basi diganti dengan data yang lebih baru dan lebih segar.
- Bergantung pada bagaimana penerapannya, perangkat penyimpanan dalam memori dapat mendukung penyimpanan *schemaless* atau *schema-aware*.
- Dukungan penyimpanan *schemaless* disediakan melalui persistensi data berbasis key-value.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Storage Devices

- Perangkat penyimpanan dalam memori sesuai jika:
 - data tiba dengan cepat dan memerlukan analisis real time atau pemrosesan *stream*
 - analitik terus menerus atau selalu aktif diperlukan, seperti BI operasional dan analitik operasional
 - pemrosesan kueri interaktif dan visualisasi data real time perlu dilakukan, termasuk analisis what-if dan operasi penelusuran
 - kumpulan data yang sama diperlukan oleh beberapa pekerjaan pemrosesan data
 - melakukan analisis data eksplorasi, karena kumpulan data yang sama tidak perlu dimuat ulang dari disk jika algoritme berubah
 - pemrosesan data melibatkan akses berulang ke kumpulan data yang sama, seperti mengeksekusi algoritme berbasis grafik
 - mengembangkan solusi Big Data latensi rendah dengan dukungan transaksi ACID



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Storage Devices

- Perangkat penyimpanan dalam memori tidak sesuai jika:
 - pemrosesan data terdiri dari pemrosesan batch
 - jumlah data yang sangat besar perlu disimpan dalam memori untuk waktu yang lama untuk melakukan analisis data yang mendalam
 - melakukan BI strategis atau analitik strategis yang melibatkan akses ke sejumlah besar data dan melibatkan pemrosesan data batch
 - kumpulan data sangat besar dan tidak sesuai dengan memori yang tersedia
 - melakukan transisi dari analisis data tradisional ke analisis Big Data, karena menggabungkan perangkat penyimpanan dalam memori mungkin memerlukan keterampilan tambahan dan melibatkan penyiapan yang rumit
 - perusahaan memiliki anggaran terbatas, karena menyiapkan perangkat penyimpanan dalam memori mungkin memerlukan peningkatan node, yang dapat dilakukan dengan penggantian node atau dengan menambahkan lebih banyak RAM



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Storage Devices

- Perangkat penyimpanan dalam memori dapat diimplementasikan sebagai:
 - In-Memory Data Grid (IMDG)
 - In-Memory Database (IMDB)



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
tinggi

QS
STARS
RATING SYSTEM
2016

ISAS
ISAS

IAEE

CPA
AUSTRALIA

ICAEW

CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA

In-Memory Data Grid (IMDG)

- IMDG menyimpan data dalam memori sebagai pasangan key-value di beberapa node di mana key dan value dapat berupa objek bisnis atau data aplikasi dalam bentuk serial.
- Ini mendukung penyimpanan data schema-less melalui penyimpanan data semi/tidak terstruktur.
- Akses data biasanya disediakan melalui API.
- Node di IMDGs tetap sinkron dan secara kolektif menyediakan ketersediaan tinggi, toleransi kesalahan, dan konsistensi.
- Dibandingkan dengan pendekatan konsistensi akhir NoSQL, IMDG mendukung konsistensi langsung.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Data Grid (IMDG)

- Dibandingkan dengan IMDB relasional, IMDG menyediakan akses data yang lebih cepat karena IMDG menyimpan data non-relasional sebagai objek.
- Oleh karena itu, tidak seperti IMDB relasional, pemetaan objek-ke-relasional tidak diperlukan dan klien dapat bekerja secara langsung dengan domain objek spesifik.
- Skala IMDGs secara horizontal dengan menerapkan partisi data dan replikasi data dan selanjutnya mendukung keandalan dengan mereplikasi data ke setidaknya satu node tambahan.
- Jika terjadi kegagalan mesin, IMDGs secara otomatis membuat ulang salinan data yang hilang dari replika sebagai bagian dari proses pemulihan.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Data Grid (IMDG)

- IMDG banyak digunakan untuk analitik real time karena mendukung Complex Event Processing (CEP) melalui model pesan publish-subscribe.
- Ini dicapai melalui fitur yang disebut kueri berkelanjutan, juga dikenal sebagai kueri aktif, di mana filter untuk peristiwa yang menarik didaftarkan ke IMDG.
- IMDG kemudian terus mengevaluasi filter dan setiap kali filter dipenuhi sebagai hasil dari operasi insert/update/delete, klien yang berlangganan diinformasikan.
- Pemberitahuan dikirim secara asinkron sebagai peristiwa perubahan, seperti added, removed dan updated, dengan informasi tentang pasangan key-value, seperti nilai lama dan baru.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Data Grid (IMDG)

- IMDG dapat ditambahkan ke solusi Big Data yang ada dengan memperkenalkannya antara perangkat penyimpanan di disk yang ada dan aplikasi pemrosesan data.
- Namun, pengenalan ini umumnya memerlukan perubahan kode aplikasi untuk mengimplementasikan API IMDGs.
- Perhatikan bahwa beberapa implementasi IMDG mungkin juga menyediakan dukungan SQL terbatas atau penuh.
- Contohnya termasuk In-Memory Data Fabric, Hazelcast, dan Oracle Coherence.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Data Grid (IMDG)

- Dalam lingkungan solusi Big Data, IMDG sering digunakan bersama dengan perangkat penyimpanan di disk yang bertindak sebagai penyimpanan backend.
- Hal ini dicapai melalui pendekatan berikut yang dapat dikombinasikan seperlunya untuk mendukung kinerja baca/tulis, konsistensi, dan persyaratan sederhana:
 - read-through
 - write-through
 - write-behind
 - refresh-ahead



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Databases

- IMDB adalah perangkat penyimpanan dalam memori yang menggunakan teknologi database dan memanfaatkan kinerja RAM untuk mengatasi masalah latensi runtime yang mengganggu perangkat penyimpanan di disk.
- IMDB dapat bersifat relasional (IMDB relasional) untuk penyimpanan data terstruktur, atau dapat memanfaatkan teknologi NoSQL (IMDB non-relasional) untuk penyimpanan data semi terstruktur dan tidak terstruktur.
- Tidak seperti IMDG, yang umumnya menyediakan akses data melalui API, IMDB relasional menggunakan bahasa SQL yang lebih dikenal, yang membantu analis data atau ilmuwan data yang tidak memiliki keterampilan pemrograman tingkat lanjut.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Databases

- IMDB berbasis NoSQL umumnya menyediakan akses berbasis API, yang mungkin sederhana menempatkan, mendapatkan, dan menghapus operasi.
- Bergantung pada implementasi yang mendasarinya, beberapa IMDB ditingkatkan, sementara yang lain ditingkatkan, untuk mencapai skalabilitas.
- Contohnya yaitu Aerospike, MemSQL, Altibase HDB, eXtreme DB dan Pivotal GemFire XD.



UNTAR
Universitas Tarumanagara



UNTAR untuk INDONESIA

In-Memory Databases

- Perangkat penyimpanan IMDB sesuai jika:
 - data relasional perlu disimpan dalam memori dengan dukungan ACID
 - menambahkan dukungan real-time ke solusi Big Data yang ada saat ini menggunakan penyimpanan di disk
 - perangkat penyimpanan pada disk yang ada dapat diganti dengan teknologi setara dalam memori
 - diperlukan untuk meminimalkan perubahan pada lapisan akses data dari kode aplikasi, seperti ketika aplikasi terdiri dari lapisan akses data berbasis SQL
 - penyimpanan relasional lebih penting daripada skalabilitas



UNTAR
Universitas Tarumanagara

Terakreditasi
BAN-PT

A
Linggi

QS
STARS
RATING SYSTEM

ISAS
KUALITAS

IAEE

CPA
AUSTRALIA

ICAEW
CHARTERED
ACCOUNTANTS

UNTAR untuk INDONESIA