

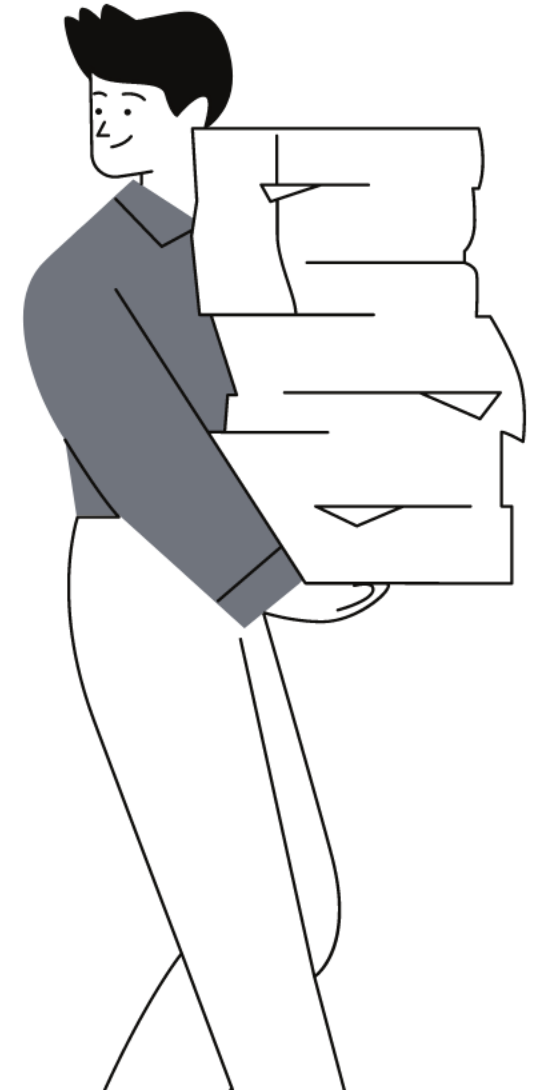


Universitas Bosowa
Teknologi Informasi

Abdillah.S.A.S.,S.Kom.,M.Pd.

Teknologi Database

Hadoop II



Perkenalan

Abdillah.S.A.S.,S.Kom.,M.Pd. bisa di panggil Pak **Abdhi** ,Bisa menjadi dosen dan teman sharing yang baik.

Lebih Lanjut Follow



Abdillah.S.A.S.,S.Kom.,M.Pd.



085242179993



Abdhi Griffindors



@abdhi



@abdhi_tux



abdhisas.com



HADOOP MapReduce

Hadoop MapReduce



Mengapa MapReduce diperlukan sejak awal hadirnya Hadoop? Dalam pendekatan tradisional, data besar diproses di master node. Ini adalah kerugian, karena dibutuhkan lebih banyak waktu untuk memproses berbagai jenis data.

Untuk mengatasi masalah ini, data diproses pada setiap slave node, dan hasil akhirnya dikirim ke master node.

MapReduce

Hadoop MapReduce



Inilah gunanya MapReduce. Di sini data diproses di mana pun data tersebut disimpan.

MapReduce didefinisikan sebagai model pemrograman di mana sejumlah besar data diproses secara paralel dan terdistribusi.

Namun, kerangka kerja MapReduce tidak bergantung pada satu bahasa tertentu. Kode program MapReduce dapat ditulis dalam Java, Python, atau Bahasa pemrograman lainnya.

MapReduce

MapReduce *task*



Seperti namanya, MapReduce terdiri dari dua tugas (task):

1. Map Tasks
2. Reduce Tasks

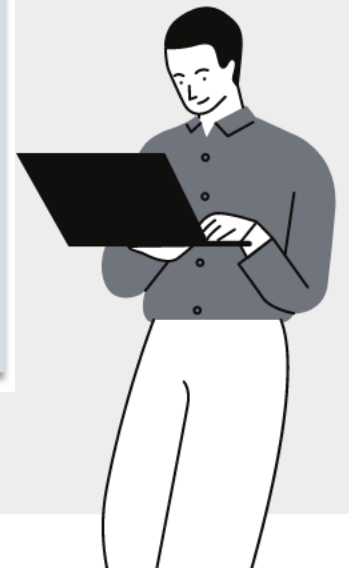
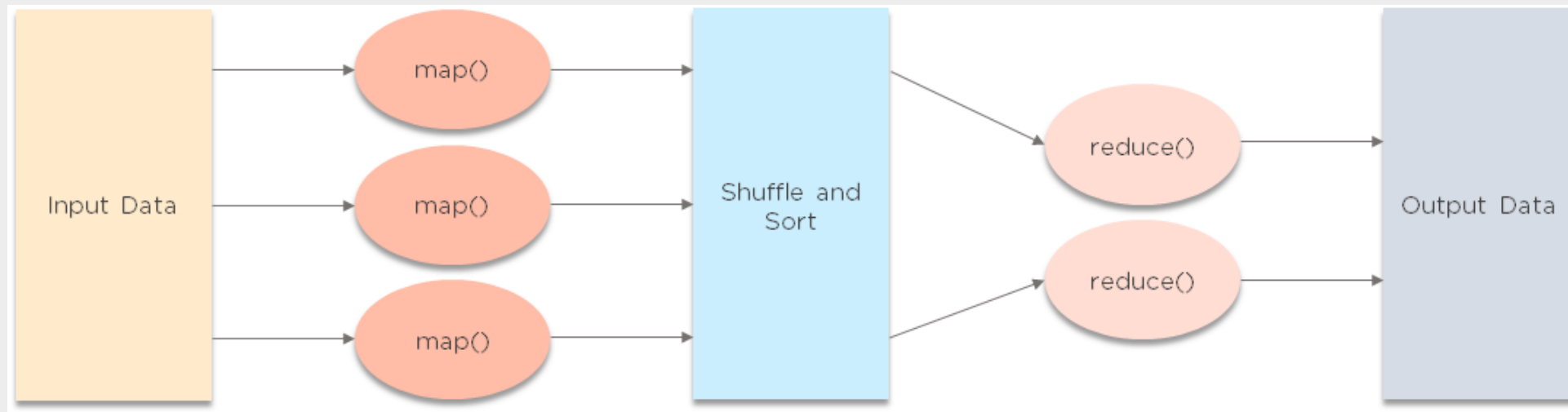


Mapper

Mapper adalah fungsi yang menangani fase pemetaan dan fungsi Reducer menangani fase reduksi.

Kedua fungsi ini menjalankan tugas Map dan Reduce secara internal.

Dari diagram berikut dapat dilihat bagaimana setiap langkah dieksekusi di dalam MapReduce.



Mapper



Pertama, data input akan dibagi (displit) menjadi sejumlah blok data. Pada fase pemetaan, fungsi mapper yang terdiri dari beberapa kode akan berjalan pada satu atau lebih (banyak) split.

Hasil dari fase pemetaan akan masuk ke fase pengocokan (shuffling) dan pengurutan (sorting), di mana output dari fase pemetaan akan dikelompokkan untuk diproses lebih lanjut.

Kemudian dalam fase pengurangan (reduce) hasilnya dikumpulkan dan nilai output tunggal dikirimkan.

Pada framework ini pengembang menyediakan tugas mapper dan reducer.

Framework MapReduce sendiri menangani shuffling, pengurutan, dan pemartisian.

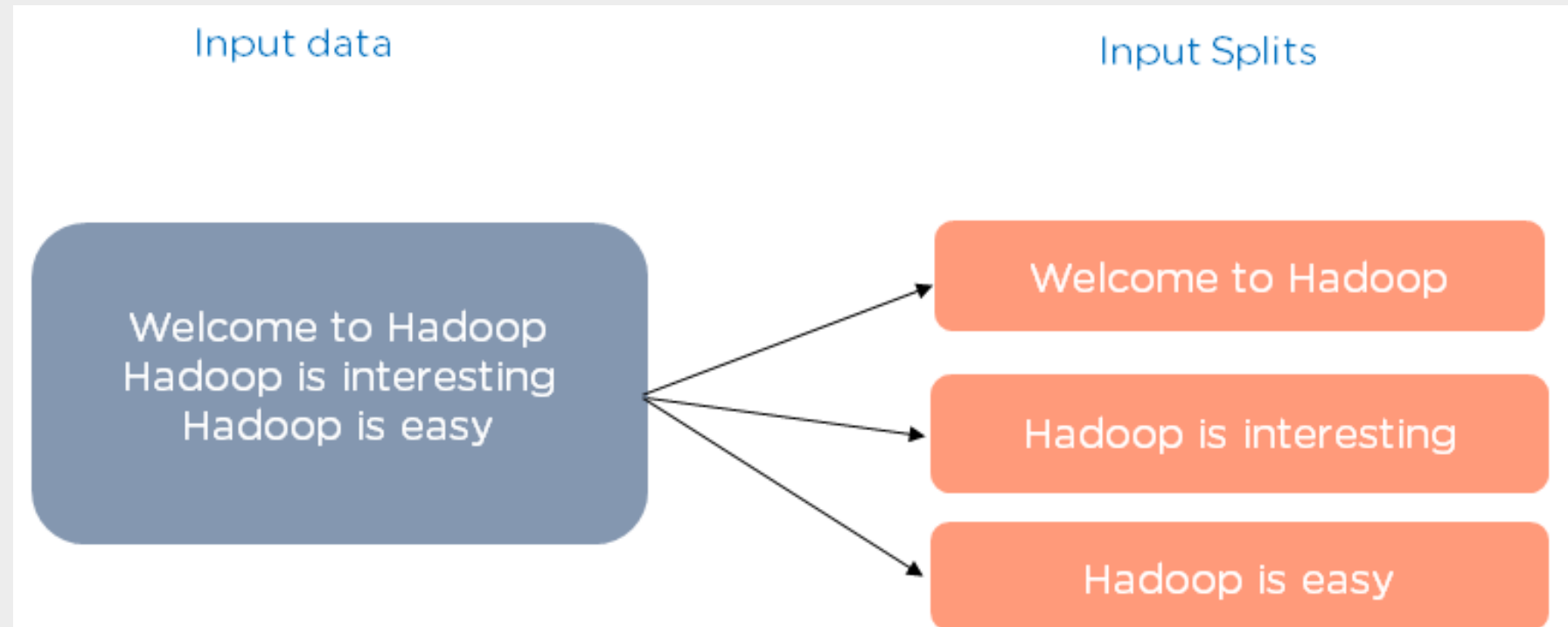
Dalam contoh berikut, data input akan dipecah, dishuffle, dan digabungkan (aggregate) untuk mendapatkan hasil akhir.



Penjelasan Proses MapReduce



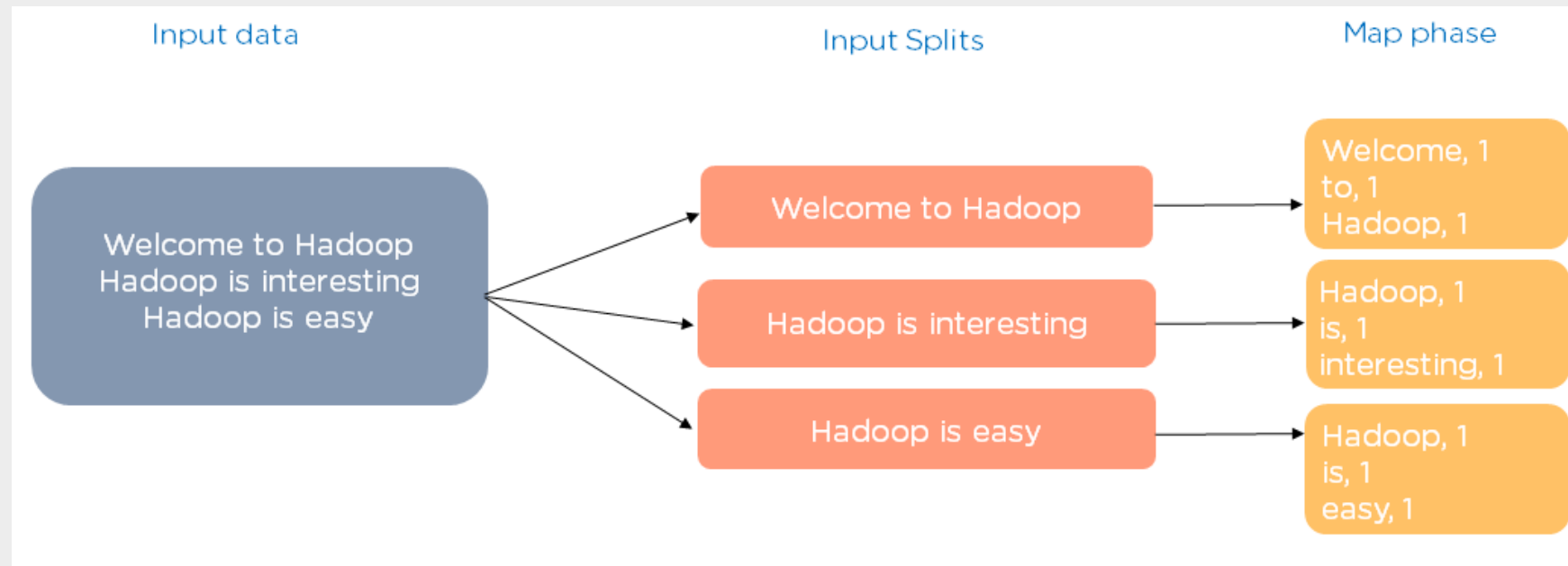
1. Data input displit, baris demi baris



Penjelasan Proses MapReduce



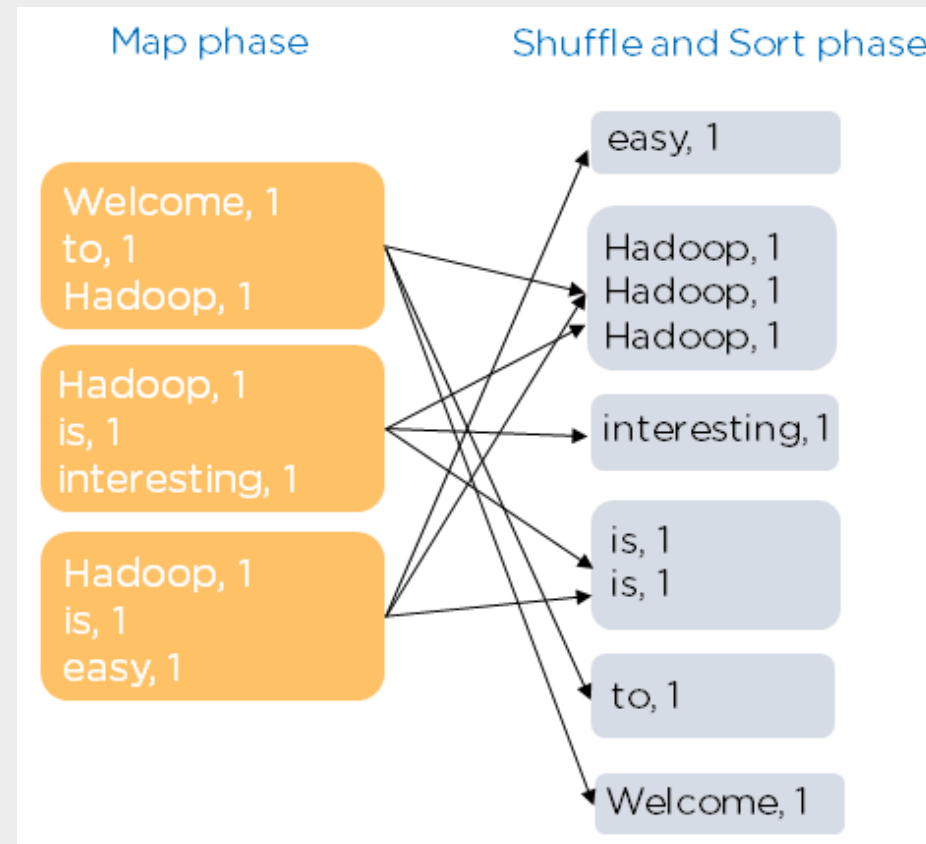
2. Fungsi *mapper* kemudian bekerja pada setiap input split yang bekerja mirip dengan model jumlah kata. Data dipetakan ke pasangan (kunci, nilai). Di sini, kita memiliki kata sebagai kunci dan nilainya sebagai satu.



Penjelasan Proses MapReduce

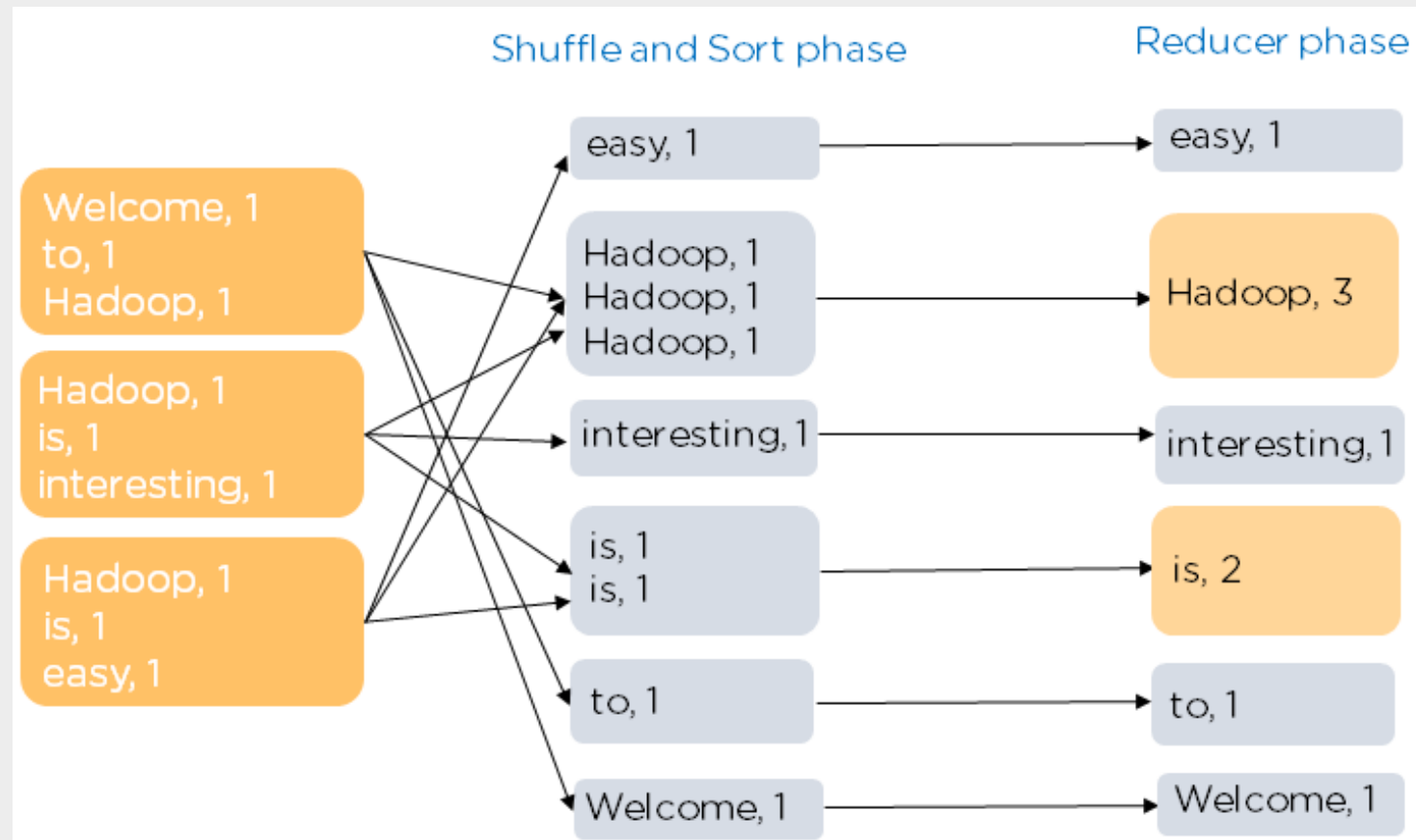


3. Pada langkah berikutnya, data dikocok dan disortir untuk mendapatkan kunci (kata) yang sama.



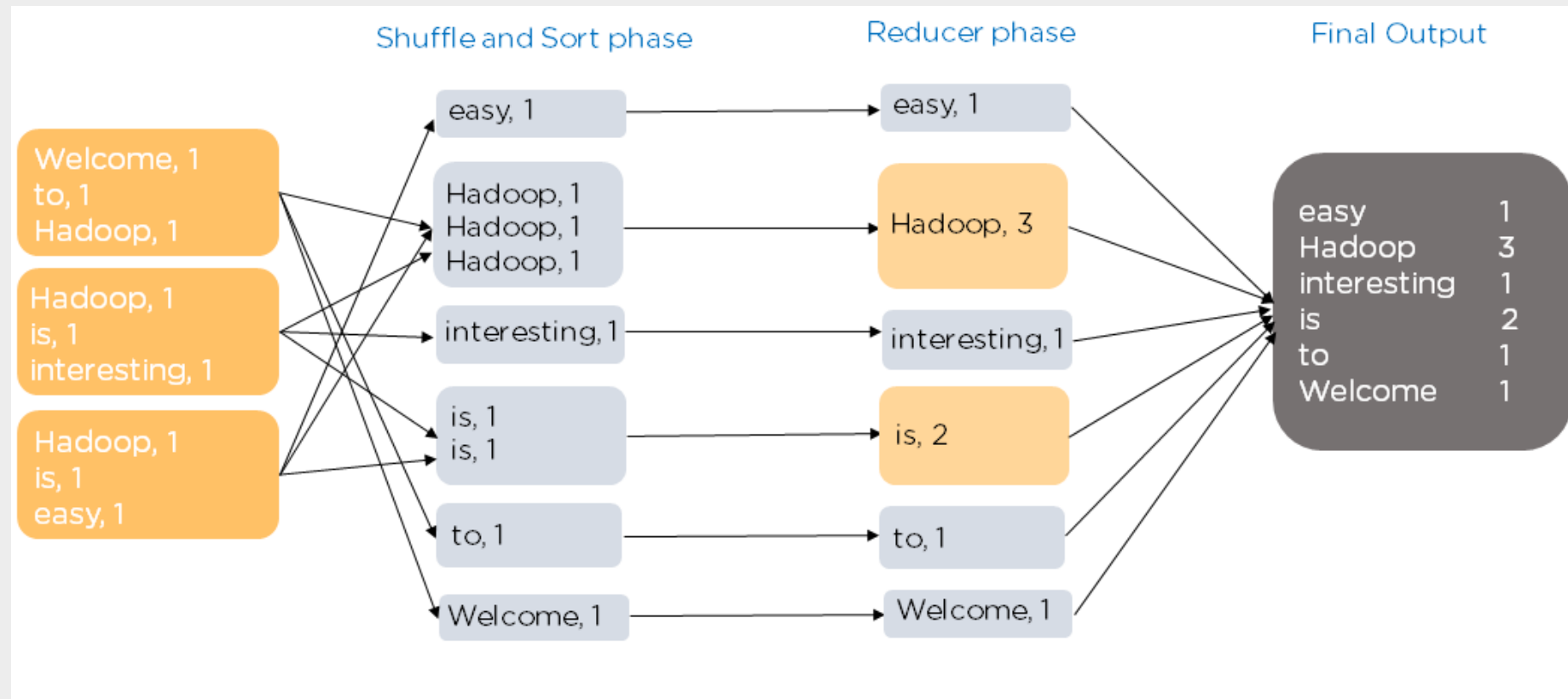
Penjelasan Proses MapReduce

4. Fase reducer mengumpulkan nilai untuk kunci yang serupa.



Penjelasan Proses MapReduce

5. Akhirnya, output akan terdiri dari daftar kata dan nilainya yang menampilkan jumlah kemunculannya.



Fitur MapReduce



Beberapa fitur MapReduce termasuk

1. Penyeimbangan muatan (load balancing) ditingkatkan karena tahapan dibagi ke dalam Map dan Reduce.
2. Ada eksekusi ulang otomatis jika tugas tertentu gagal
3. MapReduce memiliki salah satu model pemrograman paling sederhana, yang didasarkan pada Java.



Kesimpulan MapReduce



HDFS dan MapReduce adalah dua unit Hadoop 1.0. Versi ini memiliki masalah karena JobTracker melakukan pemrosesan data dan alokasi sumber daya. Hal ini mengakibatkan JobTracker menjadi terbebani.

Untuk mengatasi masalah ini, Hadoop 2.0 memperkenalkan YARN sebagai lapisan (layer) pemrosesan.



HADOOP YARN

Hadoop YARN



Di bagian sebelumnya, kita membahas bagaimana data disimpan dan diproses. Tetapi bagaimana unit alokasi sumber daya diurus? Bagaimana sumber daya dinegosiasikan di seluruh cluster? YARN menangani ini dan bertindak sebagai unit manajemen sumber daya Hadoop.

YARN (Yet Another Resource Negotiator)



YARN

Apache YARN



Apache YARN terdiri dari:

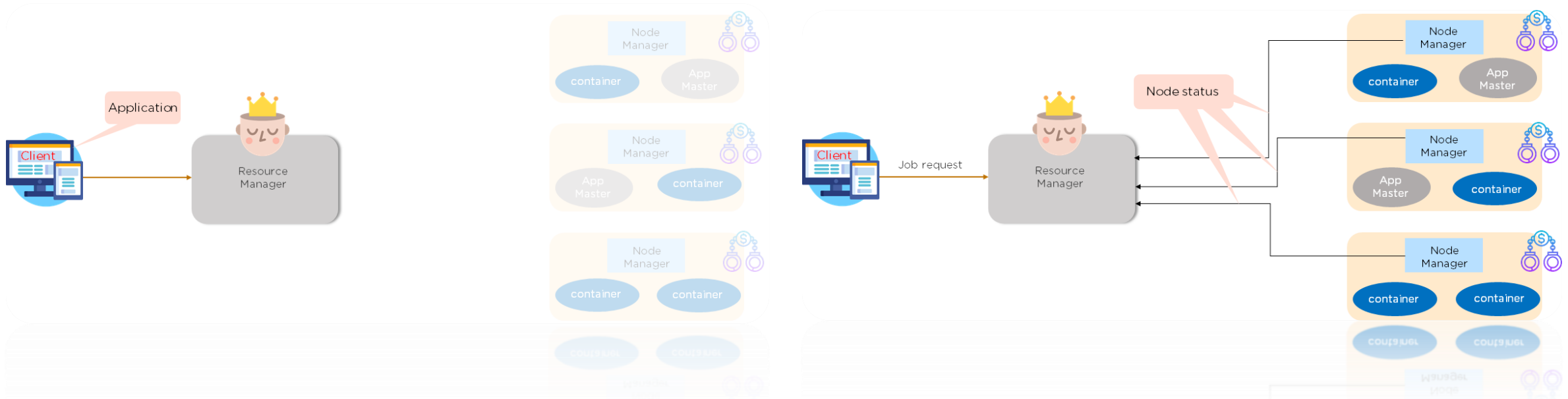
1. **Resource Manager** yang bertindak sebagai master daemon dan bertanggungjawab dalam penugasan CPU, memori, dll.
2. **Node Manager** yang bekerja sebagai slave daemon dan melaporkan pekerjaannya kepada Resource Manager.
3. **Application Master** yang bekerja dengan Resource Manager dan Node Manager dalam menegosiasikan sumber daya.



Cara Hadoop YARN Bekerja



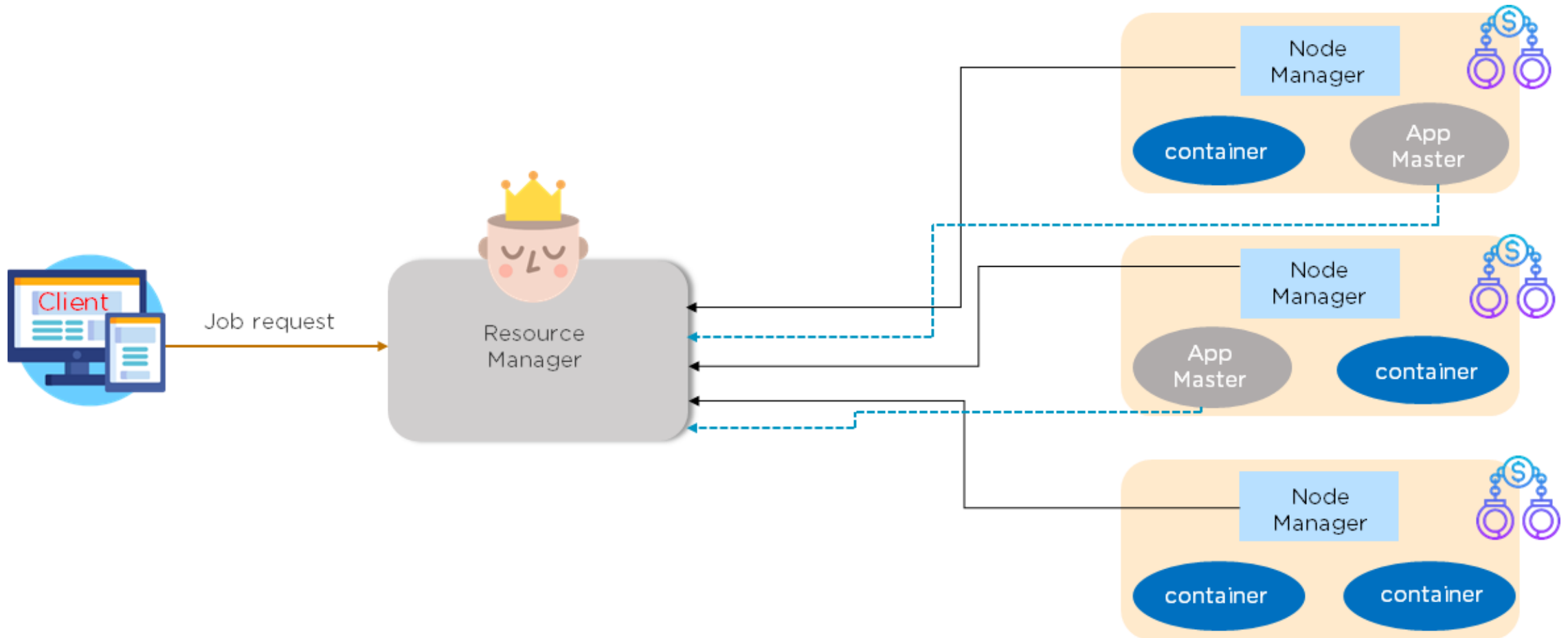
- Proses dimulai oleh klien dengan berinteraksi dengan NameNode untuk mengetahui DataNodes mana yang tersedia untuk pemrosesan data.
- Setelah langkah itu selesai, klien berinteraksi dengan Resource Manager untuk melacak sumber daya yang tersedia yang dimiliki Node Manager.
- Node Manager juga mengirim detak jantung ke Resource Manager.



Cara Hadoop YARN Bekerja



- Ketika klien menghubungi Resource Manager untuk pemrosesan data, Resource Manager ini akan meminta sumber daya yang diperlukan dari beberapa Node Manager.
- Di sini, Container adalah kumpulan sumber daya fisik seperti CPU dan RAM.
- Bergantung pada ketersediaan container ini, Node Manager memberikan respon kepada Resource Manager.
- Setelah ketersediaan dikonfirmasi, Resource Manager akan mulai menjalankan master aplikasi (App. Master)
- App. Master adalah suatu jenis kode yang mengeksekusi aplikasi.
- Ini berjalan di salah satu container dan menggunakan container lainnya untuk menjalankan tugas.
- Jika App. Master membutuhkan sumber daya tambahan, maka ia tidak dapat menghubungi Node Manager secara langsung tetapi harus tetap menghubungi Resource Manager.



Berikut ini adalah beberapa fitur YARN:

1. YARN bertanggung jawab untuk memproses permintaan pekerjaan dan mengalokasikan sumber daya
2. Versi MapReduce yang berbeda dapat berjalan di YARN sehingga upgrade MapReduce dapat dikelola
3. Bergantung pada kebutuhan, kita dapat menambahkan node.



Kesimpulan



Kita telah belajar tentang big data, Hadoop dan keterkaitan keduanya. Kita telah melihat apa dan bagaimana kerja dari HDFS, MapReduce, dan YARN.

Jika ingin mempelajari lebih lanjut tentang big data dan Hadoop, bacalah beberapa buku tentang Big data dan Hadoop



Ada Pertanyaan ?

Ayo Diskusi, dan
Belajar Bersama.

Are you ready?

Terima Kasih

