

## Abstrak

Pada sebuah perusahaan besar, sangatlah penting untuk memiliki manajemen sistem basisdata yang mampu menampung seluruh data dan dokumen milik karyawan. Data tersebut akan sangat besar sehingga tidak akan memungkinkan untuk ditampung oleh *single server*. Untuk menyelesaikan masalah tersebut, data yang sangat besar itu dapat didistribusikan ke dalam beberapa *cluster*. Di lingkungan terdistribusi inilah implementasi dari metode MapReduce akan sangat bermanfaat bagi sistem. MapReduce adalah sebuah operasi untuk menyelesaikan masalah yang mirip dengan algoritma *divide and conquer*. Sesuai dengan namanya, MapReduce terdiri dari proses *map* (pemetaan) suatu data dan *reduce* (pengurangan) yang berakhir pada penggabungan data-data yang sama.

Pada tugas akhir ini akan dilakukan penelitian terhadap performansi MapReduce di lingkungan sistem basisdata terdistribusi homogen. Untuk membangun lingkungan tersebut, sebelumnya harus dilakukan pengecekan terhadap setiap komputer yang digunakan. Pastikan bahwa seluruh komputer memiliki spesifikasi perangkat lunak dan keras, serta manajemen sistem basisdata yang sama. Setelah itu, dataset yang berbasis dokumen harus di-*import* ke *database* komputer yang berperan sebagai *master*. Kemudian, dengan menggunakan metode *sharding*, setiap node akan diberi peran: *master* akan berperan sebagai *router*, satu *node* sebagai *config server*, dan sisanya sebagai *shard server* sehingga terbentuklah lingkungan sistem basisdata berbasis dokumen terdistribusi homogen. Dataset kemudian akan didistribusikan ke setiap *shard*. Akhirnya, query MapReduce akan dijalankan dan diuji di *single server* dan 3 arsitektur *distributed database* yang berbeda untuk diteliti performansinya.

Dari hasil pengujian yang dilakukan, dapat dilihat bahwa MapReduce bekerja lebih baik di lingkungan terdistribusi dibandingkan dengan pada *single server*. Kesimpulan yang dapat diambil adalah bahwa sistem basisdata terdistribusi meningkatkan performansi MapReduce.

Kata kunci: *MapReduce, document-oriented database, distributed database system, distributed database management system, homogeneous distributed database*