

INTISARI

IMPLEMENTASI SISTEM REKOMENDASI MULTI-CRITERIA COLLABORATIVE FILTERING PADA CLUSTER APACHE SPARK

ARDHI WIJAYANTO
13/357335/PPA/04474

Skalabilitas menjadi masalah yang dihadapi oleh sistem rekomendasi yang menggunakan algoritme *collaborative filtering*. Masalah yang sama juga dihadapi ketika menggunakan algoritme *multi-criteria collaborative filtering*. Ketika ukuran dataset yang diolah semakin meningkat secara signifikan, kinerja *multi-criteria collaborative filtering* menjadi kurang maksimal. Penelitian ini berusaha untuk mengatasi masalah skalabilitas yang dihadapi *multi-criteria collaborative filtering* dengan menerapkan pendekatan *scale-out* atau menambahkan *node* ke sistem.

Proses pembangkitan rekomendasi pada sistem rekomendasi *multi-criteria collaborative filtering* akan dilakukan pada jaringan komputer multi *node* yang terdapat pada *cluster* menggunakan *framework* Apache Spark. Dataset hotel multikriteria yang digunakan untuk membuat rekomendasi disebar menjadi beberapa partisi sesuai jumlah *node* yang digunakan. Algoritme *multi-criteria collaborative filtering* diimplementasikan secara paralel menggunakan API RDD yang terdapat pada *framework* Apache Spark untuk mengolah dataset yang telah dipartisi. Untuk mengukur kinerja sistem, dinilai skalabilitas berdasarkan nilai *speedup* dari perbandingan waktu eksekusi yang didapatkan dari program rekomendasi *multi-criteria collaborative filtering* yang dijalankan pada berbagai ukuran *cluster*.

Berdasarkan hasil pengujian, didapatkan nilai *speedup* *multi-criteria collaborative filtering* pada *cluster* Apache Spark dengan 5 *node worker* adalah sebesar 2,342 yang kemudian meningkat menjadi 3,760 pada *cluster* yang mempunyai 10 *node worker*. Nilai *speedup* meningkat kembali menjadi 4,229 pada *cluster* yang memuat 15 *node worker* dan menjadi 5,564 ketika dijalankan pada *cluster* dengan 20 *node worker*.

Kata kunci : *sistem rekomendasi, multi-criteria collaborative filtering, skalabilitas, Apache Spark*

ABSTRACT

THE IMPLEMENTATION OF MULTI-CRITERIA COLLABORATIVE FILTERING RECOMMENDATION SYSTEM ON APACHE SPARK CLUSTER

ARDHI WIJAYANTO
13/357335/PPA/04474

Scalability is a problem commonly faced by a recommendation system that uses collaborative filtering methods. Multi-criteria collaborative filtering recommendation system facing the exact same problem. The performance of multi-criteria collaborative filtering is reduced when the amount of data processed by recommendation system is increasing too high. This research aim to solve the scalability problem of multi-criteria collaborative filtering algorithm by applying scale-out approach or adding computer node to run the recommendation system. The dataset will be distributed through certain number of partitions depend on the size of the cluster. Multi-criteria collaborative filtering algorithm will be implement in parallel way using RDD API of Apache Spark framework.

The process of generating a recommendation on multi-criteria collaborative filtering recommendation system will be done on a multiple nodes computer network inside a cluster using Apache Spark framework. To measure the performance of implemented system, the execution time of multi-criteria collaborative filtering recommendation system that are implemented in a recommendation program on Apache Spark will be compared in the form of speedup value.

Based on test results, it is known that as the number of size of datasets are increased. The speedup value is 2,342 on cluster with 5 worker nodes and increased to 3,760 inside a cluster with 10 worker nodes. The speedup value on cluster with 15 worker nodes become 4,229 and increased to 5,564 when use cluster with 20 worker nodes.

Keywords : *recommendation system, multi-criteria collaborative filtering, scalability, Apache Spark*