

Tugas I IF4044

Teknologi Big Data

HADOOP



KNOWLEDGE & SOFTWARE ENGINEERING

Objektif

1. Mahasiswa mampu mengoperasikan HDFS dan memahami komponen utama HDFS
2. Mahasiswa mampu melakukan implementasi MapReduce pada HDFS berdasarkan kasus yang diberikan
3. Mahasiswa mampu menjelaskan hasil eksekusi MapReduce job yang dilakukan

Problem Statement

Diberikan sekumpulan file JSON berukuran ~1GB. File-file tersebut adalah file hasil crawling dari social media meliputi Twitter, Facebook, Instagram, Youtube. Dari file-file tersebut, mahasiswa diminta untuk membuat sebuah MapReduce job yang mampu mengagregasi file-file tersebut menjadi sebuah CSV file yang memiliki kolom sebagai berikut:

- 1. social_media
- 2. date
- 3. count

Proses agregasi menggunakan social media dan timestamp yang dinormalisasi jadi perhari sebagai acuan agregasi, serta count adalah jumlah post/komentar pada dari social media yang masuk ke acuan tersebut

Delivearable

1. Dokumen sederhana yang menjelaskan stack yang digunakan dan alur implementasi MapReduce job
2. Source code yang diunggah di repository public berbasis git yang berisi implementasi MapReduce job
3. Demo eksekusi implementasi MapReduce job dari mulai bagaimana file JSON diunggah ke HDFS sampai file CSV hasil MapReduce bisa dibuka di laptop mahasiswa

Waktu Pengumpulan

1. Deliverable 1 dan 2 tanggal 1 Maret 2023 di LMS
2. Deliverable 3 tanggal 2 Maret offline di tatap muka minggu ke-7

