

ANALISIS TREN MUSIC DI SPOTIFY MENGUNAKAN HADOOP MAPREDUCE

Ade Sulastri, Alvian Ardiansyah, Alvin Rahmat Saputra,
Muhammad Akbar Perdana, Adha Mashur Sajiah

Teknik Informatika, Universitas Halu Oleo
Kampus Hijau Bumi Tridharma, Kota Kendari, Indonesia
adesulastri22001@gmail.com

ABSTRAK

Penelitian ini menganalisis tren musik di Spotify selama 10 tahun terakhir dengan memanfaatkan Hadoop MapReduce untuk memahami pola konsumsi musik di era digital. Spotify, sebagai platform streaming musik terkemuka, menghasilkan data besar yang mencerminkan preferensi pengguna, termasuk genre, artis, dan lagu yang paling populer. Tujuan penelitian ini adalah mengidentifikasi tren musik berdasarkan genre dominan, artis terpopuler, serta lagu yang paling sering didengarkan. Dataset yang digunakan diperoleh dari Kaggle dan playlist Spotify, terdiri atas 807 entri dengan tiga variabel utama: judul lagu, artis, dan genre. Proses analisis dilakukan melalui implementasi Hadoop MapReduce, meliputi tahapan pemetaan data, reduksi, hingga visualisasi hasil. Hasil penelitian menunjukkan bahwa genre pop mendominasi dengan kemunculan sebanyak 38% dari total dataset. Artis seperti The Weeknd, Justin Bieber, dan Taylor Swift tercatat sebagai artis dengan frekuensi kemunculan tertinggi. Lagu *Blinding Lights* oleh The Weeknd menjadi salah satu lagu dengan jumlah kemunculan terbanyak. Secara keseluruhan, Hadoop MapReduce berhasil mengidentifikasi tren musik secara efisien, memberikan wawasan penting bagi industri musik dalam merancang strategi distribusi dan promosi berdasarkan preferensi pasar yang terukur.

Kata kunci: Analisis Data, Big Data, Hadoop, MapReduce, Spotify, Tren Musik

1. PENDAHULUAN

Pasar musik rekaman global tumbuh sebesar 10,2% sehingga menjadikan pertumbuhan ini tertinggi kedua dalam sejarah. Nilai pasar mencapai US \$28,6 miliar, pendapatan naik di semua wilayah dan hampir semua format musik rekaman, mencerminkan industri yang benar-benar global dan beragam. Penggerak utama dalam pertumbuhan dan pangsa pasar berasal dari pendapatan streaming. Khususnya, streaming berlangganan tumbuh sebesar 11,2% dan menyumbang 48,9% dari total pasar global [1]. Salah satu *app streaming* yang paling diminati yaitu Spotify. Spotify menjadi salah satu platform yang memiliki sekitar sepertiga dari semua pendengar *streaming* di seluruh dunia dan sebagai platform pertama yang memiliki lebih dari 50 juta pelanggan [2].

Layanan streaming musik menjadi metode paling populer bagi konsumen untuk mendengarkan musik. Layanan streaming menawarkan konsumen akses tak terbatas ke katalog musik besar. Layanan streaming musik yang cukup populer yaitu Spotify, Joox, dan yang terbaru adalah YouTube Music. Sementara layanan *streaming video* antara lain YouTube, Hooq, Netflix, iFlix, Viu, dan Vidio. Layanan ini menyimpan musik di server yang dapat disambungkan pengguna melalui laptop dan perangkat seluler mereka. Sejak munculnya saluran digital untuk distribusi media industri musik telah mengalami proses transformasi besar, yang ditandai dengan turunnya pendapatan dan pertumbuhan yang kuat dalam penjualan digital dalam beberapa tahun terakhir [3].

Musik telah menjadi bagian dari kehidupan manusia, memainkan peran penting dalam kesehatan mental dan fisik manusia, identitas budaya, dan hiburan sehari-hari. Rata-rata orang mendengarkan musik 20,7 jam tercatat tahun 2023, meningkat dibanding tahun 2022. Tidak hanya para generasi muda yang sering mendengarkan musik dalam melakukan aktivitas sehari-hari, namun generasi dewasa hingga generasi tua memiliki tingkat keterlibatan signifikan dalam mendengarkan musik sebagai bagian dari rutinitas sehari-hari, seperti bekerja, bersantai atau bahkan dalam olahraga [4].

Dengan meningkatnya minat manusia dalam mendengarkan musik dan Spotify menjadi *app streaming* pertama yang memiliki banyak pelanggan untuk mendengarkan musik, sehingga menghasilkan data besar yang mencerminkan referensi, tren, dan pola perilaku pendengar dalam mengkonsumsi musik. Penelitian ini bertujuan untuk menganalisis tren musik di Spotify selama sepuluh tahun terakhir menggunakan Hadoop, sehingga dapat mengetahui popularitas genre, artis yang terpopuler selama sepuluh tahun terakhir, dan musik yang menjadi konsumsi paling banyak di Spotify.

Pemilihan Hadoop MapReduce dalam penelitian ini didasarkan pada kemampuannya dalam memproses data dalam skala besar secara terdistribusi [5]. Mengingat data lagu dari Spotify bersifat masif dan terus bertambah, MapReduce menawarkan efisiensi waktu dan skalabilitas dalam pemrosesan data, serta sangat sesuai digunakan dalam lingkungan big data seperti proyek ini.

2. TINJAUAN PUSTAKA

2.1 Big Data

Analisis Big Data adalah suatu fenomena yang terjadi secara fundamental dan mampu mengubah apa yang kita lakukan dan ketahui. Big data analytics menjelaskan bagaimana entitas dapat memperoleh, menyimpan, membagi, melakukan evaluasi, dan mengerjakan suatu kegiatan didasarkan informasi yang diciptakan oleh perangkat elektronik dan manusia sebagai pengguna, yang nantinya akan didistribusikan melalui komputer dan teknologi jaringan. Pada dasarnya sistem big data analytics sendiri mengacu pada 3 hal utama yaitu volume, variety, velocity (3V) dan unsur V lain yaitu veracity dan value yang erat kaitannya dengan karakteristik sebuah data yang relevan baik secara kepastian serta manfaat. Dalam organisasi yang menjadikan aplikasi computer menjadi sumber informasi utama dalam menghasilkan data, sedikit banyak akan meningkatkan jumlah volume data secara eksponensial [6].

2.2 Hadoop



Gambar 1. Logo aplikasi hadoop

Hadoop adalah salah satu *platform opensource* yang dapat digunakan untuk yang dapat digunakan untuk menyimpan, memproses, mengatur, dan menganalisis data dalam jumlah besar dari jenis data terstruktur, semi terstruktur dan tidak terstruktur. Hadoop menggunakan model komputasi Mapreduce, yang memungkinkan aplikasi dipecah menjadi sejumlah aplikasi yang lebih kecil yang dapat dijalankan oleh satu node dalam cluster. Selain itu, Hadoop menyediakan sistem file yang didistribusikan yang disebut *Hadoop Distributed File System* (HDFS) yang menyebarkan data ke seluruh node dalam cluster. Platform ini dirancang untuk menawarkan ketersediaan data dan pergerakan aplikasi serta *bandwith cluster* yang tinggi [7].

2.3 Hadoop Distributed File System (HDFS)

Hadoop distributed file system merupakan sistem penyimpanan terdistribusi, yang dapat melakukan pemecahan pada file besar menjadi bagian-bagian kecil selanjutnya dilakukan disebarkan ke setiap cluster dari komputer. HDFS terdiri dari dua komponen utama yaitu NameNode dan DataNode. NameNode berfungsi sebagai komputer utama (*master*) dalam Hadoop Cluster sedangkan DataNode adalah komputer-komputer dalam cluster yang berperan sebagai *slaves* atau pendukung. NameNode bertanggung jawab untuk menyimpan informasi tentang lokasi blok-blok data dalam cluster, serta mengatur dan mengawasi persebaran blok-blok data di seluruh DataNode. Sementara DataNode bertanggung jawab untuk menyimpan blok-blok data yang telah dialokasikan kepadanya, dan secara teratur

melaporkan kepada NameNode tentang *update* status dan kondisinya [8].

2.4 Mapreduce

MapReduce digunakan untuk mengolah *big data* dalam sistem file Hadoop dengan membagi data menjadi bagian-bagian kecil kemudian menyatukan kembali hasilnya. Mapreduce terdiri dari dua tahap yaitu *map* dan *reduce*. Pada tahap *map* berfungsi untuk membaca inputan dalam format yang memiliki pasangan *key/value*, sementara untuk tahap *reduce* berfungsi untuk membaca format *key/value* yang telah diinputkan kemudian dikelompokkan berdasarkan key yang sama [9].

2.5 Yarn

YARN (Yet Another Resource Negotiator) adalah resource manager multiguna yang dapat menjadwalkan dan mengatur siklus CPU dari cluster Hadoop ke aplikasi yang akan dijalankan. YARN merupakan peningkatan dari MapReduce yang dikenalkan pada Hadoop versi 2.0 yang dikenal juga sebagai MapReduce 2.0. Dalam Hadoop YARN dideskripsikan sebagai alat yang memungkinkan framework pemrosesan data lain dijalankan di Hadoop, sehingga Hadoop tidak hanya dapat menjalankan MapReduce [10].

2.6 Java Development Kit (JDK)

JDK (Java Development Kit) adalah produk dari oracle corporation yang ditujukan untuk pengembang bahasa java. jdk sebenarnya merupakan software development kit yang khusus dikeluarkan oleh oracle. jdk menggunakan lisensi gnu public license (gpl) yang menjadikan software yang free dan open source [11].

2.7 Jobs

Job merupakan suatu pekerjaan yang diberikan kepada Hadoop dan dapat dikerjakan oleh Hadoop. Job dalam kategorisasinya dapat dikelompokkan menjadi homogen (job yang sama) atau dilakukan secara bersamaan antar job yang homogen (job yang berbeda) [12].

2.8 Spotify

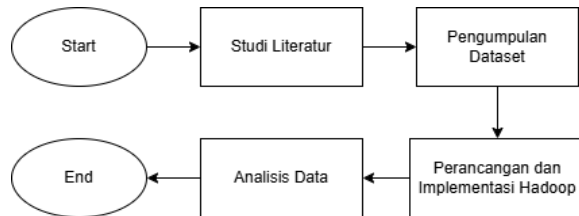


Gambar 2. Logo aplikasi spotify

Spotify adalah aplikasi yang diperuntukkan mendengarkan banyak lagu dan podcast secara streaming (Spotify, 2023). Aplikasi music streaming ini memiliki akses freemium yang mana pengguna dapat mengakses secara gratis dengan beberapa keterbatasan. Aplikasi asal swedia ini sudah diunduh lebih dari 100 ribu pengguna. Berdasarkan data dari Data Books (2022) terdapat 456 juta orang yang menggunakan aplikasi tersebut pada kuartal II. Angka

tersebut meningkat sebanyak 5,31% dibandingkan tahun sebelumnya. Dari indikasi tersebut, dapat dikatakan bahwa spotify sebagai salah satu platform yang cukup digemari [13].

3. METODE PENELITIAN



Gambar 3. Diagram alur metode penelitian

3.1. Studi Literatur

Studi kasus yang diangkat dalam percobaan ini bertujuan untuk menganalisis tren musik sepanjang

periode 13 tahun, mulai dari tahun 2010 hingga 2023. Dalam rangka mencapai tujuan tersebut, kami melakukan tinjauan literatur yang mendalam guna memperoleh pemahaman yang komprehensif mengenai dasar-dasar Framework Hadoop.

3.2. Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini bersumber dari Kaggle, serta diekstraksi langsung dari playlist terpopuler di Spotify untuk mencakup rentang waktu antara tahun 2010 hingga 2023 yang mana setiap playlist terdiri dari 50 daftar lagu. Dataset ini awalnya terdiri dari 15 variabel, namun penelitian difokuskan pada tiga variabel utama yang dianggap paling relevan, yaitu judul lagu, nama artis, dan genre musik. Jumlah total data yang digunakan mencapai 807 entri dengan format file CSV, yang dipilih untuk mempermudah proses analisis dan pengolahan data.

Tabel 1. Sampel data

Lagu	Artis	Genre	Tahun
Hey, Soul Sister	Train	neo mellow	2010
Love The Way You Lie	Eminem	detroit hip hop	2010
TiK ToK	Kesha	dance pop	2010
Bad Romance	Lady Gaga	dance pop	2010
Just the Way You Are	Bruno Mars	pop	2010
...	...	...	...
Tyler The Creator	See You Again (feat. Kali Uchis)	['hip hop', 'rap']	2023
Me Porto Bonito	Bad Bunny	['reggaeton', 'trap latino', 'urbano latino']	2023
OMG	NewJeans	['k-pop', 'k-pop girl group']	2023
AMG	Natanael Cano	['corrido', 'corridos tumbados', 'musica mexic...']	2023
Romantic Homicide	d4vd	['bedroom pop']	2023

3.3. Perancangan dan Implementasi Hadoop

Sebelum mengimplementasikan Hadoop Framework untuk analisis data musik di Spotify, langkah pertama yang harus dilakukan adalah merencanakan proses implementasi secara terperinci. Perencanaan ini mencakup instalasi Hadoop pada master node dan slave node, serta konfigurasi setiap *node*, termasuk pengaturan *bashrc*, *namenode*, *datanode*, dan parameter terkait lainnya, untuk memastikan komunikasi yang efisien antar node dengan dukungan Java. Setelah tahapan instalasi dan konfigurasi selesai dengan benar, Hadoop dan HDFS akan siap berfungsi secara optimal, memungkinkan analisis data musik Spotify dengan efisien dan efektif [14].

Berikut implementasi Hadoop untuk menganalisis data:

a. Instalasi Hadoop dan Menjalankan hadoop

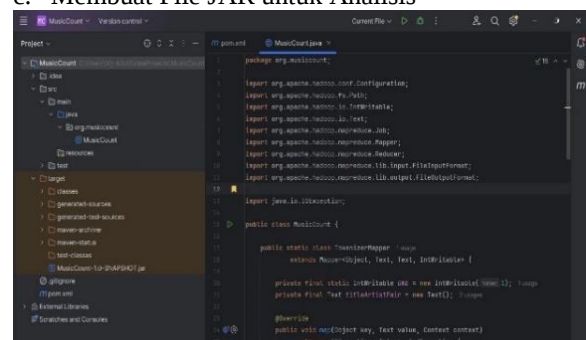
Tahap awal instalasi Hadoop mencakup persiapan lingkungan sistem dan konfigurasi perangkat lunak. Sistem operasi yang digunakan harus berbasis Linux atau Windows dengan dukungan subsistem Linux, disertai instalasi *Java Development Kit* (JDK) yang kompatibel sebagai prasyarat utama. Perangkat lunak Hadoop diunduh melalui situs resmi Apache, kemudian diekstraksi dan diikuti dengan konfigurasi file inti, yaitu *core-*

site.xml, *hdfs-site.xml*, *mapred-site.xml*, dan *yarn-site.xml*.

b. Input Data ke HDFS

Membuat direktori baru di HDFS menggunakan perintah `hdfs dfs -mkdir /input_dir`. Setelah direktori berhasil dibuat, dataset diunggah ke dalam direktori tersebut menggunakan perintah `hdfs dfs -put datasetMusic.csv /input_dir`.

c. Membuat File JAR untuk Analisis



Gambar 4. Tampilan aplikasi IntelliJ IDE

Tahap selanjutnya adalah pembuatan file JAR untuk analisis data. Proses ini dimulai dengan penulisan kode MapReduce dalam bahasa Java yang dirancang untuk mengolah dataset.

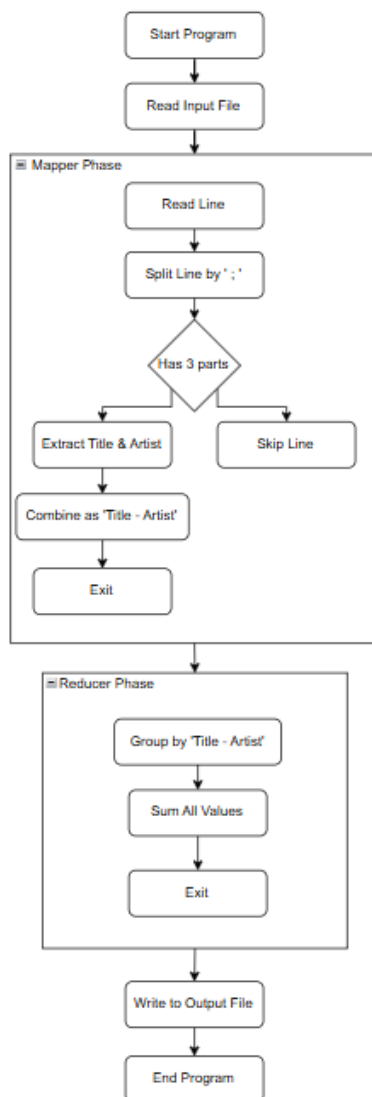
d. Menjalankan MapReduce

Tahap ke-empat adalah menjalankan MapReduce, di mana file JAR yang telah dibuat digunakan untuk menjalankan proses analisis. Proses ini dapat dipantau melalui *web interface* Hadoop di `http://localhost:8088` untuk melihat status job.

e. Output

Setelah proses selesai, hasil analisis disimpan di direktori `/output_dir` pada HDFS. Jika diperlukan, hasil analisis ini dapat diunduh ke sistem lokal menggunakan perintah `hdfs dfs -get /output_dir /lokasi_lokal`.

3.4. Analisis Data



Gambar 5. Tampilan perancangan sistem

Tahap Analisis Data dimulai dengan mengambil output hasil MapReduce dari *Hadoop Distributed File System* (HDFS) ke sistem lokal untuk diproses lebih lanjut. Data yang dihasilkan seringkali memerlukan pembersihan, seperti penghapusan duplikasi atau penyesuaian format menjadi struktur yang lebih mudah dianalisis, misalnya CSV. Setelah data bersih, langkah berikutnya adalah visualisasi untuk

mempermudah interpretasi, menggunakan alat seperti Matplotlib atau Seaborn untuk membuat grafik tren, diagram batang, atau pie chart, yang menggambarkan pola genre musik berdasarkan tahun. Dari visualisasi ini, pola atau wawasan penting dapat diidentifikasi, seperti genre paling populer, tren musiman, atau potensi prediksi popularitas di masa depan.

Gambar 5 menggambarkan alur kerja sebuah program berbasis MapReduce untuk memproses data dari sebuah file input. Program dimulai dengan membaca file input, kemudian masuk ke Tahap Mapper. Pada tahap ini, setiap baris data dibaca dan dipisahkan menjadi beberapa bagian menggunakan tanda pemisah (`;`). Jika baris memiliki tiga bagian, program akan mengekstrak informasi berupa judul dan artis, lalu menggabungkannya dalam format "Judul - Artis". Baris yang tidak sesuai format akan dilewati. Setelah proses pemetaan selesai, data diteruskan ke Tahap Reducer, di mana data yang telah dipetakan dikelompokkan berdasarkan kombinasi "Judul - Artis". Untuk setiap grup, nilai-nilai yang relevan dijumlahkan, misalnya untuk menghitung total kemunculan. Hasil akhir dari proses ini ditulis ke dalam file output, dan program berakhir. Alur ini menunjukkan bagaimana data diproses secara sistematis untuk menghasilkan informasi yang terstruktur dan bermanfaat.

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Dataset

Dataset ini terdiri dari 807 entri yang mencakup informasi judul lagu, nama artis, genre utama, dan tahun rilis. Data mencakup lagu-lagu populer dari tahun 2010 hingga 2023, yang sebagian besar bersumber dari Kaggle dan daftar putar Spotify. Meski terdapat 105 nilai kosong pada kolom genre, kolom ini tetap menyediakan informasi utama tentang genre musik, sementara kolom judul dan artis mencantumkan nama lagu serta artis secara berurutan. Kolom tahun memberikan data kronologis yang komprehensif terkait tren musik selama satu dekade terakhir.

Dataset ini menawarkan peluang analisis terhadap tren musik, seperti distribusi genre, dominasi artis tertentu, serta dinamika popularitas genre dalam periode waktu tertentu. Kehadiran data genre, meski tidak sepenuhnya lengkap, tetap relevan untuk mengidentifikasi pola utama dengan pendekatan analitik berbasis big data, seperti Hadoop MapReduce. Dengan kombinasi sumber data dari Kaggle dan Spotify, analisis ini bertujuan untuk memberikan wawasan mendalam mengenai evolusi tren musik, karakteristik genre dominan, serta perilaku pendengar di platform streaming digital.

4.2. Implementasi Hadoop MapReduce

a. Jalankan Aplikasi Hadoop

Pertama, pastikan Hadoop sudah terinstall dengan benar serta file JDK juga terinstall dengan benar. Jika sudah, jalankan perintah `start-all.cmd` pada

terminal untuk menjalankan Hadoop-nya. Setelah itu, nonaktifkan safe mode agar kita bisa menambah dan menghapus file pada HDFS

```
C:\Windows\System32>start-all.cmd
This script is Deprecated. Instead use start-hadoop.cmd

C:\Windows\System32>cd\

C:\>hadoop dfsadmin -safemode leave
DEPRECATED: Use of this script to execute hdfs commands is deprecated.
Instead use the hdfs command for it.
Safe mode is OFF
```

Gambar 6. Perintah menjalankan hadoop

b. Buat Direktori Input dan Masukan File ke HDFS

```
C:\>hadoop fs -mkdir /input_dir

C:\>hadoop fs -put C:/dataset/Music.csv /input_dir

C:\>hadoop fs -ls /input_dir/
Found 1 items
-rw-r--r-- 1 Alvian supergroup 29685 2024-12-10 18:48 /input_dir/Music.csv
```

Gambar 7. Perintah buat dir dan input file

Untuk menyiapkan direktori input di HDFS, jalankan perintah *hadoop filesystem make directory input_dir* untuk membuat direktori input. Setelah itu, unggah dataset ke direktori tersebut dengan perintah *hadoop filesystem put dataset_path input_dir*. Terakhir, verifikasi keberadaan dataset dengan menjalankan perintah *hadoop filesystem list input_dir*.

c. Jalankan File JAR dan Buat Direktori Output

```
C:\>hadoop jar C:/MusicCount-1.0-SNAPSHOT.jar org.musiccount.MusicCount /input_dir /output_dir
2024-12-10 18:50:38,630 INFO client.DefaultHadoopFailoverProxyProvider: Connecting to ResourceManagers
2024-12-10 18:50:44,343 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing
2024-12-10 18:50:44,349 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path:
27400216.0001
2024-12-10 18:50:44,621 INFO input.FileInputFormat: Total input files to process : 1
2024-12-10 18:50:44,686 INFO mapreduce.JobSubmitter: number of splits:1
2024-12-10 18:50:44,803 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1733827400216_0001
2024-12-10 18:50:44,803 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-12-10 18:50:44,971 INFO conf.Configuration: resource-types.xml not found
2024-12-10 18:50:44,972 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-12-10 18:50:45,404 INFO impl.YarnClientImpl: Submitted application application_1733827400216_0001
2024-12-10 18:50:45,449 INFO mapreduce.Job: The url to track the job: http://Alvian-Ardiansyah:8080/jobdetails/job_1733827400216_0001
2024-12-10 18:50:45,451 INFO mapreduce.Job: Running job: job_1733827400216_0001
2024-12-10 18:51:18,870 INFO mapreduce.Job: Job job_1733827400216_0001 running in uber mode : false
2024-12-10 18:51:18,871 INFO mapreduce.Job: map 0% reduce 0%
```

Gambar 8. Perintah menjalankan file JAR

File JAR yang dibuat digunakan untuk menjalankan tugas MapReduce, yang bertujuan menghitung jumlah kemunculan kombinasi judul lagu dan artis berdasarkan kumpulan data masukan. Tiga

komponen utama dalam program ini adalah Mapper, Reducer, dan Driver. Mapper memproses data dalam bentuk baris, memetakan setiap kombinasi judul dan artis sebagai pasangan kunci-nilai (judul-artis, 1). Combiner digunakan untuk mengoptimalkan proses dengan menggabungkan data sementara sebelum diteruskan ke Reducer. Reducer kemudian mengakumulasi data untuk menghasilkan jumlah total kemunculan tiap kombinasi judul dan artis.

Hasil akhir adalah daftar judul lagu dan artis serta jumlah kemunculannya, yang disimpan dalam berkas keluaran. Driver mengatur dan menjalankan proses MapReduce dengan mengonfigurasi kelas Mapper, Reducer, tipe data, kunci, nilai, serta parameter input dan output sesuai argumen eksekusi. Proses ini memungkinkan analisis efisien atas data musik dalam skala besar.

d. Simpan Hasil Analisis

```
C:\>hadoop dfs -cat /output_dir/*
DEPRECATED: Use of this script to execute hdfs commands is deprecated.
Instead use the hdfs command for it.
"CAN'T STOP THE FEELING! (Original Song from DreamWorks Animation's "How to Train Your Dragon 2") - Justin Bieber"
"Celebrate (From the Original Motion Picture "Fifty Shades of Grey") - From The "Fifty Shades of Grey" Soundtrack"
1
"Happy - From "Despicable Me 2"" - Pharrell Williams
"How Far I'll Go - From "Moana"" - Alessia Cara
"Just Like Fire (From the Original Motion Picture "The Hunger Games: Mockingjay - Part 1") - DNCE
"Let It Go - From "Frozen / Single Version" - Idina Menzel
"Love Me Like You Do - From "Fifty Shades of Grey" Soundtrack"
1
```

Gambar 9. Perintah menyimpan hasil analisis

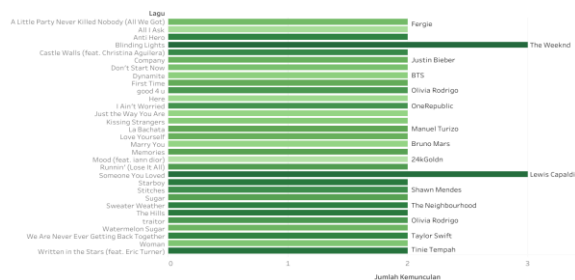
Hasil analisis di atas menunjukkan output dari program Hadoop MapReduce yang menghitung jumlah kemunculan kombinasi judul lagu dan artis berdasarkan dataset input. Setiap baris output berisi kombinasi "judul lagu - artis" diikuti dengan angka yang menunjukkan jumlah kemunculannya. Sebagian besar lagu memiliki frekuensi kemunculan sebesar 1, menandakan bahwa lagu-lagu ini unik dalam dataset. Namun, ada beberapa kasus seperti lagu "A Little Party Never Killed Nobody (All We Got)" oleh Fergie yang muncul sebanyak 2 kali, menunjukkan bahwa lagu tersebut terdaftar lebih dari sekali dalam dataset. Output ini memberikan informasi dasar tentang frekuensi lagu dan artis tertentu yang dapat digunakan untuk analisis lebih lanjut, seperti mendeteksi duplikasi data atau mengidentifikasi tren berdasarkan popularitas lagu.

4.3. Hasil Analisis Tren Musik

Tabel 2. Hasil analisis

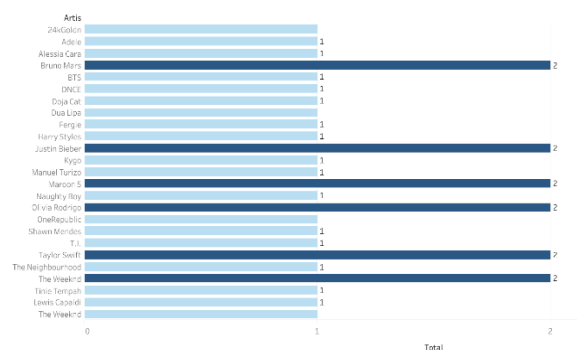
Lagu	Artis	Jumlah Kemunculan
Blinking Lights	The Weeknd	3
Someone You Loved	Lewis Capaldi	3
A Little Party Never Killed Nobody (All We Got)	Fergie	2
All I Ask	Adele	2
Anti Hero	Taylor Swift	2
Castle Walls (feat. Christina Aguilera)	T.I.	2
Company	Justin Bieber	2
Don't Start Now	Dua Lipa	2

Lagu	Artis	Jumlah Kemunculan
Dynamite	BTS	2
First Time	Kygo	2
Here	Alessia Cara	2
I Ain't Worried	OneRepublic	2
Just the Way You Are	Bruno Mars	2
Kissing Strangers	DNCE	2
La Bachata	Manuel Turizo	2
Love Yourself	Justin Bieber	2
Marry You	Bruno Mars	2
Memories	Maroon 5	2
Mood (feat. iann dior)	24kGoldn	2
Runnin' (Lose It All)	Naughty Boy	2
Starboy	The Weeknd	2
Stitches	Shawn Mendes	2
Sugar	Maroon 5	2
Sweater Weather	The Neighbourhood	2
The Hills	The Weeknd	2
Watermelon Sugar	Harry Styles	2
We Are Never Ever Getting Back Together	Taylor Swift	2
Woman	Doja Cat	2
Written in the Stars (feat. Eric Turner)	Tinie Tempah	2
good 4 u	Olivia Rodrigo	2
traitor	Olivia Rodrigo	2



Gambar 10. Grafik visualisasi kemunculan lagu

Hasil analisis pada Gambar 10 menunjukkan bahwa Hadoop MapReduce dapat mengidentifikasi lagu dengan jumlah kemunculan terbanyak dari dataset Spotify dalam rentang waktu 2010 hingga 2023. lagu *Blinding Lights* oleh *The Weeknd* dan *Someone You Loved* oleh *Lewis Capaldi* mendominasi hasil analisis dengan masing-masing mencatatkan tiga kali kemunculan dalam rentang waktu 2010 hingga 2023. Dengan kata lain, lagu-lagu ini berhasil bertahan dalam daftar lagu populer selama tiga tahun, yang merupakan durasi maksimal kemunculan lagu-lagu populer dalam rentang waktu analisis.



Gambar 11. Hasil analisis kemunculan artis

Hasil analisis pada Gambar 11 memperlihatkan visualisasi hasil analisis terhadap frekuensi kemunculan artis dalam dataset Spotify. Grafik ini membantu memberikan gambaran umum tren popularitas artis dari waktu ke waktu. Sumbu X menunjukkan nama artis, sedangkan sumbu Y menunjukkan jumlah kemunculan lagu dari masing-masing artis.

Berdasarkan grafik tersebut menunjukkan bahwa artis Bruno Mars, Justin Bieber, Maroon 5, Olivia Rodrigo, Taylor Swift, dan The Weeknd masing-masing tercatat muncul sebanyak dua kali berdasarkan jumlah lagu populer yang mereka hasilkan dalam periode 2010 hingga 2023.

5. KESIMPULAN DAN SARAN

Penelitian ini membuktikan bahwa Hadoop MapReduce efektif digunakan dalam menganalisis data besar seperti dataset lagu terhits di Spotify dalam rentang waktu 13 tahun. Proses MapReduce mampu memetakan dan mereduksi data lagu serta artis dengan hasil yang akurat. Namun, keterbatasan dalam jumlah node dan konfigurasi sistem masih menjadi tantangan. Sehingga dapat disimpulkan dari analisis tersebut bahwa tren musik di Spotify dipengaruhi oleh popularitas artis, genre, dan daya tarik lintas waktu dari lagu-lagu tertentu. Untuk lagu terpopuler paling lama bertahan selama 3 tahun karena kemunculan lagu yang paling banyak yaitu sebanyak 3 kali sedangkan untuk artis yang populer hanya bisa bertahan selama 2 tahun karena kemunculan terbanyak hanya 2 kali. Penelitian selanjutnya dapat mengembangkan sistem dengan menambahkan jumlah node untuk meningkatkan performa dan efisiensi. Selain itu, dapat dilakukan eksplorasi terhadap framework lain seperti Apache Spark atau melakukan analisis sentimen

terhadap lirik lagu guna memberikan perspektif tren yang lebih dalam terhadap preferensi musik pengguna.

DAFTAR PUSTAKA

- [1] “from IFPI. Use of this report, IFPI Global Music Report (2024 edition) is governed by the Global Music Report Terms of Use (gmr.ifpi.org/terms-of-use) and Global Music Report Content Usage Rules (gmr.ifpi.org/content-usage). Designed By Data Design,” 2023.
- [2] David Curry, “Music Streaming App Revenue and Usage Statistics (2024).”
- [3] D. Noviani, R. Pratiwi, S. Silvianadewi, M. B. Alexandri, and M. A. Hakim, “Pengaruh Streaming Musik Terhadap Industri Musik Di Indonesia,” 2020.
- [4] “Engaging with music 2 0 2 3.”
- [5] M. D. Khairy, A. N. Defitri, A. Hadi, A. Nuzuli, A. L. Dyakiyyah, and A. Heryanto, “NetPLG Journal of Network and Computer Applications Pengolahan Data Sensor IOT Dengan Apache Spark Menggunakan Metode Batch Processing”, [Online]. Available: <https://jurnal.netplg.com/jnca>
- [6] H. Arifulsyah, H. Ribut Yuliantoro, and dan Abdi Bhayangkara, “Pengaruh Penerapan Analisis Big Data Dalam Pengambilan Keputusan Investasi Saham,” 2023. [Online]. Available: <https://jurnal.pcr.ac.id/index.php/jakb/>
- [7] K. Junaidi, A. C. Aurelie, S. Israwana, A. Putra, and A. Ryu, “Implementasi MapReduce Pada Dataset Spotify Top Music Untuk Mengetahui Artist yang paling banyak Didengar Dalam Kurun Waktu 10 Tahun,” vol. 2, no. 2, 2023, [Online]. Available: <https://jurnal.netplg.com/>
- [8] L. Arrahmando Romadhona, R. Febrianti, A. Winata, C. Putri Amanda, and R. Julia Erizka, “NetPLG Journal of Network and Computer Applications Hadoop-MapReduce Pada YARN Framework,” *NetPLG Journal of Network and Computer Applications*, vol. Vol. 1, pp. 91–101, 2022, [Online]. Available: <https://jurnal.netplg.com/jnca>
- [9] I. Permatahati, M. N. Perdana, N. Apriadi, T. P. Amanda, and Z. Maharani, “Analisis Dataset TOP 1000 IMDb Movies Menggunakan Hadoop,” vol. 2, no. 2, 2023, [Online]. Available: <https://jurnal.netplg.com/>
- [10] C. Khairrun Nissa, A. Krishnawan, A. Budi Negoro, Z. Amira, and Z. Amru Abdillah, “NetPLG Journal of Network and Computer Applications Implementasi Ketersediaan Elemen Hadoop dan Analisis Berdasarkan Daemon Log Monitoring”, [Online]. Available: <https://jurnal.netplg.com/jnca>
- [11] D. Taruna *et al.*, “Aplikasi Pengenalan Dan Pencegahan Bencana Kebakaran Api Yang Disebabkan Oleh Manusia (Human Error) Berbasis Android,” 2021.
- [12] S. Prabowo and M. Abdurrohman, “Studi Perbandingan Performa Algoritma Penjadwalan untuk Real Time Data Twitter pada Hadoop,” *Komputika : Jurnal Sistem Komputer*, vol. 9, no. 1, pp. 43–50, Apr. 2020, doi: 10.34010/komputika.v9i1.2848.
- [13] R. U. Fahmi, A. Ghani Zahran, and S. P. Suwandi, “Analisis User Experience Terhadap Tingkat Kepuasan Pengguna Pada Aplikasi Spotify Dengan Metode Utaut,” *Jurnal Teknologi Informasi*, vol. 4, no. 1, 2023, doi: 10.46576/djtechno.
- [14] I. G. Putri, Z. Hanifa, M. Haryani, F. Pradika, and A. Aritonang, “NetPLG Journal of Network and Computer Applications Implementasi Hadoop Framework: Analisis Data Nasabah Bank Dengan Metode Mapreduce”, [Online]. Available: <https://jurnal.netplg.com/jnca>