

BAB II KAJIAN LITERATUR

2.1 Tinjauan Pustaka

Pada bagian ini penulis akan menjelaskan terkait dasaran teori yang digunakan untuk membuat sistem visualisasi dashboard untuk memonitoring Hadoop dengan Grafana.

2.1.1 Big Data

Big Data merupakan sebuah istilah yang diartikan sebagai adanya sekumpulan data yang begitu besar dan kompleks dimana tidak bisa lagi ditangani oleh teknologi konvensional atau *Database* biasa. Pada umumnya, penggunaan parameter ukuran data lebih umum digunakan untuk menjawab apakah merupakan *Big Data* atau bukan[4]. Berikut tabel perbedaan antara teknologi konvensional dengan big data:

Tabel 2. 1 Perbedaan Database Konvensional dengan Big Data

	Database Konvensional	<i>Big Data</i>
Ukuran	GB	TB atau PB
Laju Data	Perjam atau Perhari	Lebih cepat
Struktur Data	Terstruktur	Tidak terstruktur
Sumber Data	Terpusat	Tersebar
Integrasi Data	Mudah	Sulit
Penyimpanan Data	RDBMS	HDFS, NoSQL

Parameter ukuran bukanlah satu-satunya hal yang menjadi karakteristik dari *Big Data* dibagi menjadi 4 bagian atau sering disebut sebagai 4V, yaitu *Volume* (ukuran), *Variety*(variasi data), *Velocity*(kecepatan), dan juga *Veracity*. Berikut penjelasan lebih lanjut terkait karakteristik *Big Data*:



Gambar 2.1 Karakteristik Big Data

1. Volume

Merujuk pada besarnya ukuran besarnya data yang akan digunakan atau diolah baik dalam ukuran MB, GB, TB, PB, ZB dan seterusnya. Sebagai batasan minimum *Big Data* masih beragam. Namun banyak yang mengatakan bahwa *Terabytes* sebagai ukuran minimum *Big Data*.

2. Variety

Merupakan tingkat keberagaman data dalam dataset, dapat berupa struktur seperti *Database*, data semi terstruktur seperti *Json* dan *Xml*, ataupun tidak struktur yang biasanya lebih mendominasi seperti sosial media.

3. Velocity

Merupakan tingkat keberagaman data dalam dataset, dapat berupa struktur seperti *Database*, data semi terstruktur seperti *Json* dan *Xml*, ataupun tidak struktur yang biasanya lebih mendominasi seperti sosial media.

4. Veracity

Merujuk kepada keabsahan data atau kepastian kondisi data, apakah data bisa dinyatakan sebagai benar ataupun salah, sehingga *Big Data* dapat menghasilkan data yang berkualitas dan juga terpercaya.

Meskipun sudah banyak kajian big data namun masih ada beberapa tantangan dalam pemanfaatan big data sebagai berikut[5]:

1. Validitas, tidak mudah untuk melakukan uji validasi pada big data karena sifatnya besar dan heterogen
2. Akses terhadap sumber data, permasalahan yang sering terjadi terkait kerahasiaan data, oleh karena itu perlu sebuah mekanisme yang dapat menjembatani pemilik data agar tidak menimbulkan kekhawatiran
3. Metode dan pemrosesan data yang rumit

Beberapa peluang dalam pemanfaatan big data antara lain sebagai berikut:

1. Pengganti data lengkap, menggantikan survei.
2. Mengedit, membantu deteksi dan penanganan anomali pada data.
3. Menghubungkan ke data yang lain, membuat kumpulan data yang lebih kaya.

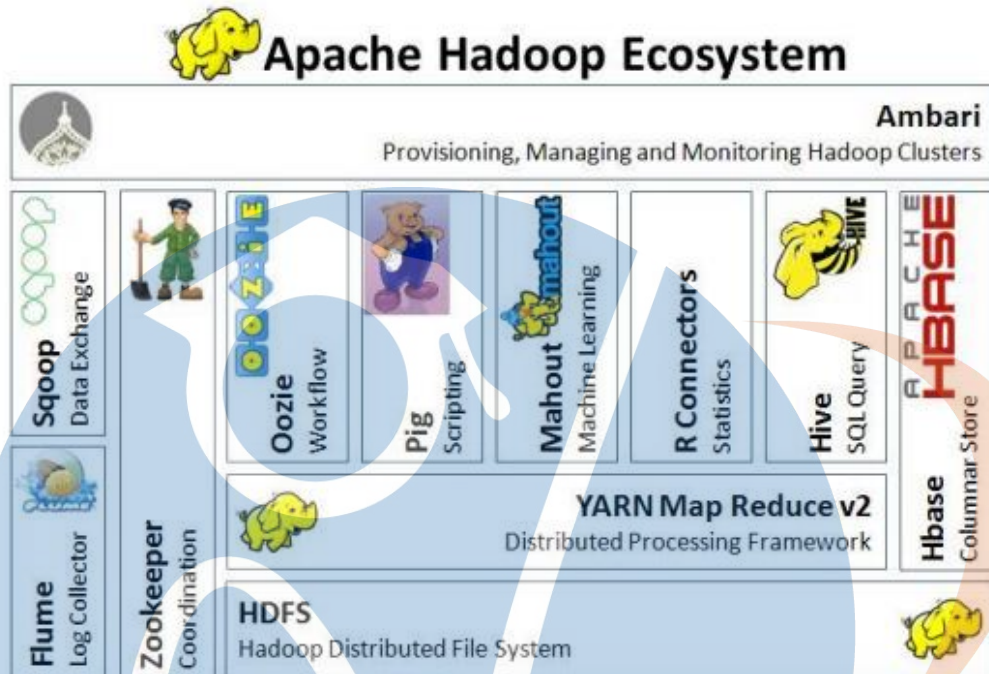
Beberapa teknologi pada big data adalah sebagai berikut:

1. Apache Hadoop
2. Apache Hive
3. Apache Spark

2.1.2 Hadoop

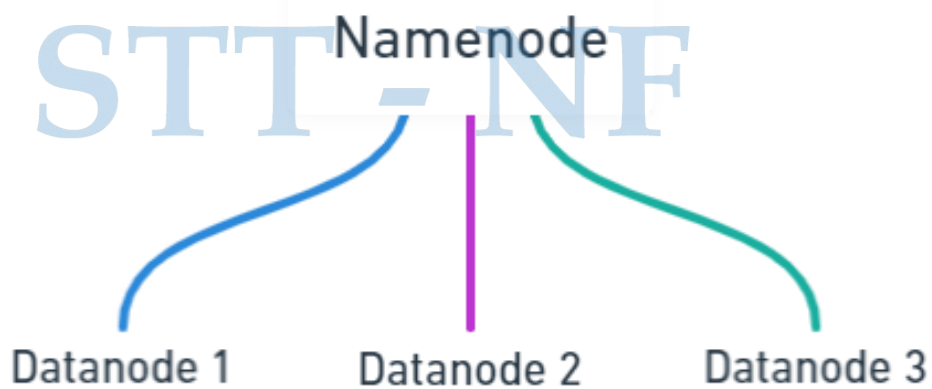
Hadoop merupakan suatu platform *Big Data* untuk sistem distribusi, yang mencakup file sistem terdistribusi dan komputasi terdistribusi yang berjalan pada suatu sistem [6]. pertama kali dikembangkan oleh Doug Cutting dan Mike Cafarella pada tahun 2006. Hadoop dapat mengidentifikasi kegunaan dan bagaimana hubungan antara atau masing-masing alat *Big Data*, dan

kebutuhan apa saja yang nantinya akan digunakan dalam pembuatan implementasi pada ekosistem hadoop.



Gambar 2. 2 Ekosistem Hadoop

Fokus dari dibuatnya visualisasi *Big Data* adalah untuk menggambarkan sistem yang berinteraksi antara tool big data ataupun lingkungannya dan digunakan untuk merancang sebuah sistem dari gabungan beberapa tool yang sesuai dengan seberapa kompleks sistem yang akan dibangun [7].



Gambar 2. 3 Arsitektur Hadoop

Berikut ini beberapa fitur utama hadoop:

1. Erasure coding

Fitur ini memberikan penghematan overhead penyimpanan sampai dengan 50%.

2. Yarn federation

Memungkinkan untuk memajemen banyak cluster menjadi satu cluster.

3. Namenode high Availability

Memungkinkan untuk memiliki lebih dari 1 namenode yang standby.

2.1.3 Monitoring Sistem

Monitoring memungkinkan sebuah basis data dapat dimonitor menggunakan tools yang bisa memberikan kita visibilitas terhadap kondisi data sehingga kita bisa melakukan aksi perbaikan dengan cepat jika terdapat masalah pada data tersebut[8].

Keunggulan dari monitoring itu sendiri adalah untuk menurunkan *downtime* yang berarti dapat menurunkan potensi kerugian, dan juga dapat dengan cepat memahami sebuah informasi terkait performa suatu sistem yang sedang dimonitoring.

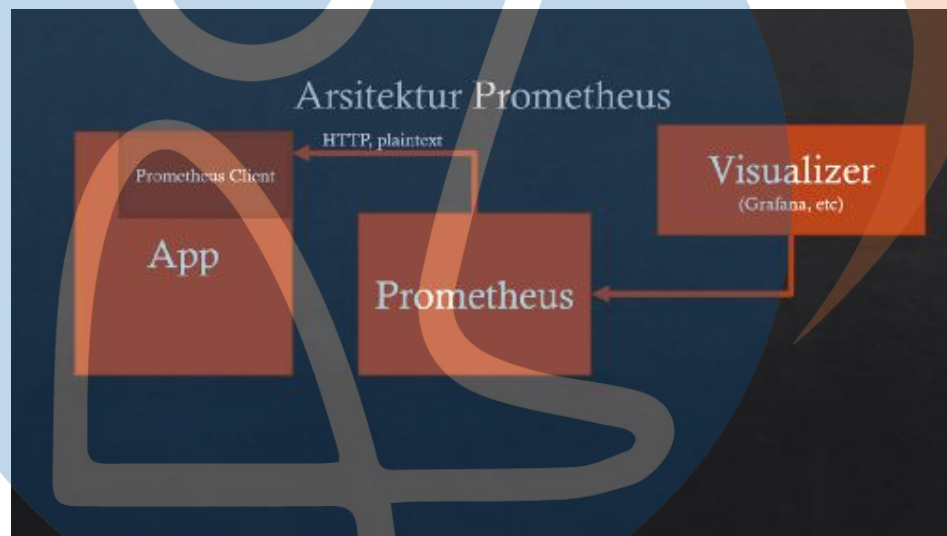
SNMP(*Simple Network Management protocol*) adalah suatu protokol yang digunakan untuk memonitoring dan mengolah perangkat jaringan. SNMP menggunakan sistem log yang menyimpan segala informasi dari perangkat jaringan yang terdaftar. Keuntungan dari SNMP sendiri adalah SNMP dapat mengambil segala kebutuhan informasi perangkat yang dibutuhkan dan dapat menampilkan untuk memenuhi kebutuhan sistem monitoring[9].

Monitoring menggunakan protokol SNMP dapat menampilkan daftar perangkat yang dipantau dan dapat melihat kondisi atau status perangkat yang di monitoring seperti *Hard Disk usage*, *RAM usage*, dan *CPU usage*. Untuk memudahkan administrator jaringan sistem menyimpan data pada waktu tertentu untuk memudahkan proses menganalisis jaringan dan Monitoring[10].

2.1.4 Prometheus

Prometheus merupakan sebuah perangkat lunak berbasis *Open Source* yang digunakan untuk melakukan monitoring dan juga alerting. Prometheus dapat dikostumisasi dengan mudah sehingga membantu saat diintegrasikan dengan perangkat lunak lainnya.

Prometheus memiliki beberapa komponen utama seperti prometheus Server yang berfungsi untuk penyimpanan dan pengelolaan data. Prometheus server melakukan pengambilan data dengan interval waktu yang ditentukan pada konfigurasi dari target tertentu dan menyimpannya dalam bentuk *Time series*[8]. Prometheus juga dapat mengirimkan data dari target kepada Prometheus Server.



Gambar 2. 4 Arsitektur Prometheus

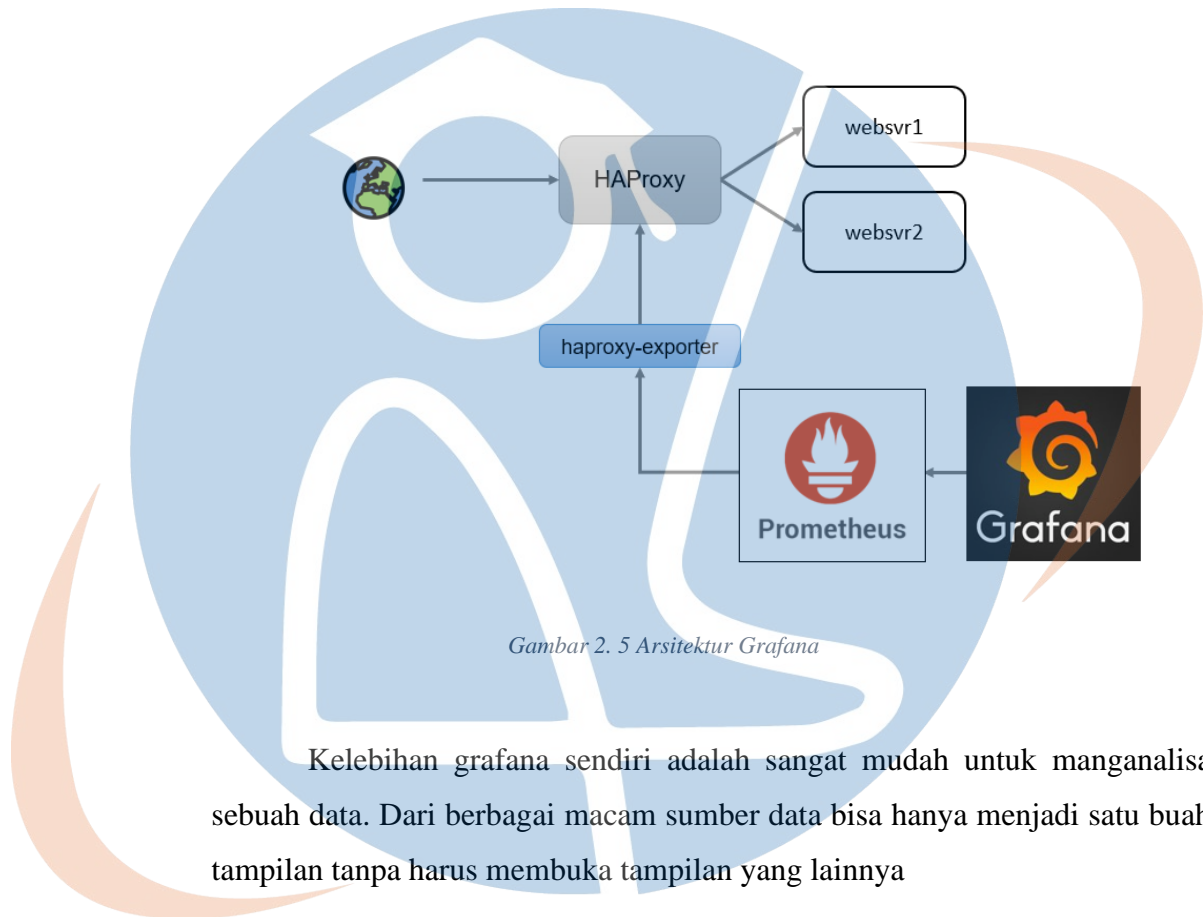
Keunggulan prometheus antara lain:

1. Mendukung multi backend sesuai dengan kebutuhan
2. Prometheus memproses pengukuran dengan protokol teks biasa
3. Prometheus akan mengambil data dari client lalu menyimpannya

2.1.5 Grafana

Grafana merupakan suatu perangkat lunak yang dapat membaca suatu data dan merubahnya kedalam bentuk grafik atau sebuah data tertulis. grafana mendukung data berbentuk *Time Series* dari sumber data, dan sering digunakan sebagai alat analisis dan monitoring [11].

Setiap sumber data memiliki *query Editor* tertentu yang disesuaikan dengan kebutuhan fitur dan juga kemampuan tertentu. Grafana mendukung banyak sumber data, salah satunya Prometheus dan juga Grafana memungkinkan untuk menggunakan *plugin* untuk diintegrasikan dengan beberapa basis data yang berbeda.



Gambar 2. 5 Arsitektur Grafana

Kelebihan grafana sendiri adalah sangat mudah untuk menganalisa sebuah data. Dari berbagai macam sumber data bisa hanya menjadi satu buah tampilan tanpa harus membuka tampilan yang lainnya

2.2 Penelitian Terkait

Pada tahap ini akan menjelaskan penelitian terkait monitoring menggunakan Prometheus dan juga Grafana. Berikut tabel yang akan menjelaskan tentang penelitian terkait:

Tabel 2. 2 Penelitian Terkait

Judul Penelitian	Peneliti	Kesimpulan
Monitoring Server dengan Prometheus dan Grafana serta Notifikasi Telegram	(Dede Rahman, Hidra Amnur, Indri Rahmayuni. 2020) (Politeknik Negeri Padang)	Monitoring server berhasil dilakukan dengan menggunakan prometheus dan grafana terhadap server. Monitoring server bekerja dengan baik sesuai harapan. Sistem operasi yang digunakan pada monitoring server adalah Ubuntu server dengan versi 18.04.sistem yang dibangun memberikan notifikasi atau pemberitahuan kepada admin apabila CPU, memori, ataupun service Apache dan MYSQL ada yang mati
Deploying and Monitoring Hadoop MapReduce Analytics on Single-chip Cloud Computer	(Andreas-Lazaros Georgiadis, Sotirios Xydis, Dimitrios Soudris. 2016) (National Technical University of Athens)	pada studi tentang perilaku runtime dari beban kerja analitik data yang muncul, yang diimplementasikan dengan kerangka kerja Hadoop MapReduce, yang diterapkan pada inti Intel SCC. Kami mempresentasikan kerangka kerja yang dikembangkan yang memungkinkan penyebaran dan pemantauan klaster Hadoop di Intel SCC sehubungan dengan keterbatasan fisik platform.
Evaluasi penggunaan Prometheus dan Grafana untuk Monitoring Database Mongoddb	(Ramadoni, Mahmud Zunus Amirudin, Rifki Fahmi, Ema Utami, Muhammad Syukri Mustafa. 2021) (Universitas Amikom Yogyakarta)	Berdasarkan keseluruhan proses analisis yang telah dilakukan oleh penulis, dapat diambil kesimpulan bahwa monitoring basis data Mongoddb akan sangat membantu administrator dalam memonitoring kinerja basis data. Dengan menggunakan Prometheus dan Grafana maka pembacaan hasil monitoring menjadi lebih mudah dan akurat
Implementasi Sistem Monitoring Menggunakan Prometheus dan Grafana	(Rahayu Mutiara Febriana. 2020) (Politeknik Negeri Jakarta)	Dari hasil penelitian yang dilakukan dapat diperoleh sebuah kesimpulan bahwa adanya sistem monitoring jaringan menggunakan Prometheus dan Grafana, dapat membantu sistem administrator jaringan untuk mengetahui kondisi jaringan yang ada. Obyek yang berjalan seperti, interface yang aktif, penggunaan CPU, penggunaan Memory dan network traffic serta jumlah perangkat yang terhubung dalam jaringan.
Implementasi dan analisis hadoop element availability berdasarkan daemon log monitoring,	(Nurahdiatna Subagya, Adityas Wijajarto, Ahmad Almaari. 2021) (Universitas Telkom Bandung)	Dengan implementasi monitoring log menggunakan log4j dapat memantau bagaimana Hadoop element bekerja ketika sistem Hadoop sedang melakukan proses. Availability dari Hadoop Core system dapat dipantau dari daemon log saat menjalankan sebuah proses

2.3 Posisi Penelitian

Pada tahap ini akan menjelaskan tentang posisi penelitian yang membedakan studi kasus dashboard monitoring menggunakan Prometheus dan juga Grafana. Berikut tabel yang akan menjelaskan perbedaan posisi penelitian yang sedang dilakukan dengan penelitian yang telah dilakukan.

Tabel 2. 3 Posisi Penelitian

No.	Hadoop	Prometheus	Grafana
1	-----	Dede Rahman, Hidra Amnur, Indri Rahmayuni. 2020 (Politeknik Negeri Padang) Monitoring Server dengan Prometheus dan Grafana serta Notifikasi Telegram	
2	Andreas-Lazaros Georgiadis, Sotirios Xydis, Dimitrios Soudris. 2016 (National Technical University of Athens) Deploying and Monitoring Hadoop MapReduce Analytics on Single-chip Cloud Computer		
3	-----	Ramadoni, Mahmud Zulus Amirudin, Rifki Fahmi, Ema Utami, Muhammad Syukri Mustafa. 2021 (Universitas Amikom Yogyakarta) Evaluasi penggunaan Prometheus dan Grafana untuk Monitoring Database MongoDB	
4	Nurahdiatna Subagya, Adityas Wijajarto, Ahmad Almaari. 2021 (Universitas Telkom Bandung) Implementasi dan analisis hadoop element availability berdasarkan daemon log monitoring,		
5	-----	Rahayu Mutiara Febriana. 2020 (Politeknik Negeri Jakarta) Implementasi Sistem Monitoring Menggunakan Prometheus dan Grafana	
6	Muhammad Akbar. 2022 (STT Terpadu Nurul Fikri) Perancangan dan implementasi Dashboard Monitoring Hadoop menggunakan Grafana.		

Berdasarkan kelima penelitian diatas, peneliti menyimpulkan bahwa terdapat kedekatan pada masing-masing penelitian yang telah dilakukan. Pada penelitian nomor satu, tiga, dan lima memiliki kedekatan dengan penelitian yang dilakukan yaitu pada bagian penggunaan *software* yang sama yaitu Prometheus dan Grafana. Sedangkan pada penelitian nomor dua dan empat memiliki persamaan pada penggunaan Hadoop. Namun pada penelitian yang akan dilakukan ini akan terfokus terhadap monitoring Hadoop dengan menggunakan *software* Prometheus untuk mengumpulkan data metrik yang dimiliki oleh Hadoop dan menggunakan Grafana sebagai dashboard monitoring Hadoop.



STT - NF