

Penjelasan:

1. Zookeeper

ZooKeeper merupakan layanan terpusat untuk memelihara informasi konfigurasi, penamaan, koordinasi kluster, menyediakan sinkronisasi terdistribusi, dan menyediakan layanan grup kluster.

2. Ambari

Ambari digunakan untuk manajemen Hadoop lebih sederhana dengan mengembangkan perangkat lunak untuk menyediakan, mengelola, dan memantau kluster Apache Hadoop. Ambari menyediakan UI web manajemen Hadoop yang intuitif dan mudah digunakan yang didukung oleh RESTful API.

3. Spark

Spark adalah mesin analitik terpadu untuk pemrosesan data skala besar. Spark menyediakan API tingkat tinggi dalam bahasa pemrograman Scala, Java, Python, dan R. Spark mendukung pemrosesan tingkat tinggi seperti Spark SQL, DataFrames, MLlib dan GraphX.

4. Kafka

Apache Kafka merupakan platform terdistribusi untuk data streaming. Pada dasarnya, Apache Kafka merupakan sistem publish/subscribe messaging, dimana terdapat satu atau lebih sistem yang men-generate data untuk suatu topik tertentu secara real-time di Apache Kafka (disebut sebagai Producers). Kemudian, topik tersebut dapat dibaca oleh satu atau lebih sistem yang membutuhkan data-data dari topik tersebut secara real-time (disebut sebagai Consumers).

5. HBase

Berfungsi sebagai non-relational distributed database, yang artinya data yang disimpan dalam HBase tidak disimpan dalam bentuk tabel. Terinspirasi dari Google's Big Table, dan seperti Hadoop yang lain ditulis dengan bahasa pemrograman Java. Biasanya dalam Hadoop ecosystem peran dari HBase adalah suporting HDFS dan posisinya berada di atas HDFS. HBase sangat berguna ketika user membutuhkan akses *read/write* yang realtime dan random ke big data.

6. NiFi

Apache NiFi merupakan alat (tool) grafik perutean data, transformasi, dan logika mediasi sistem yang kuat dan dapat diskalakan. Apache NiFi digunakan untuk mentransformasikan data sesuai dengan kebutuhan

7. Hadoop YARN

YARN "Yet Another Resource Negotiator" adalah lapisan manajemen sumber daya Hadoop. YARN diperkenalkan di Hadoop 2.x. YARN memungkinkan berbagai mesin pengolah data seperti pemrosesan grafik, pemrosesan interaktif, pemrosesan arus serta pemrosesan batch untuk menjalankan dan memproses data yang tersimpan dalam HDFS (Hadoop Distributed File System). Selain pengelolaan sumber daya, YARN juga digunakan untuk penjadwalan pekerjaan. YARN memperluas kekuatan Hadoop ke teknologi berkembang lainnya, sehingga mereka dapat memanfaatkan kelebihan HDFS (sistem penyimpanan yang paling handal dan populer di planet ini) dan kluster ekonomi.

YARN terdiri dari tiga komponen inti:

- a. ResourceManager (satu per kluster)
- b. ApplicationMaster (satu per aplikasi)
- c. NodeManager (satu per node)

8. Hadoop HDFS

HDFS menggunakan arsitektur master / slave dimana master terdiri dari satu namaNode yang mengelola metadata sistem file dan satu atau beberapa slave DataNodes yang menyimpan data aktual. File dalam namespace HDFS dibagi menjadi beberapa blok dan blok tersebut disimpan dalam satu set DataNodes. NameNode menentukan pemetaan blok ke DataNodes. DataNodes menangani operasi baca dan tulis dengan sistem file. Mereka juga mengurus pembuatan blok, penghapusan dan replikasi berdasarkan instruksi yang diberikan oleh NameNode. HDFS menyediakan shell seperti sistem file lainnya dan daftar perintah tersedia untuk berinteraksi dengan sistem file. Perintah shell ini akan dibahas dalam bab terpisah beserta contoh yang tepat.

9. Elasticsearch

Elasticsearch adalah salah satu database yang masuk ke dunia NoSQL dengan fokus di search engine database. Elasticsearch ditenagai oleh Apache Lucene yang juga merupakan search engine database yang memiliki query low level. Elasticsearch memiliki query yang lebih mudah untuk digunakan karena berbasis RESTful

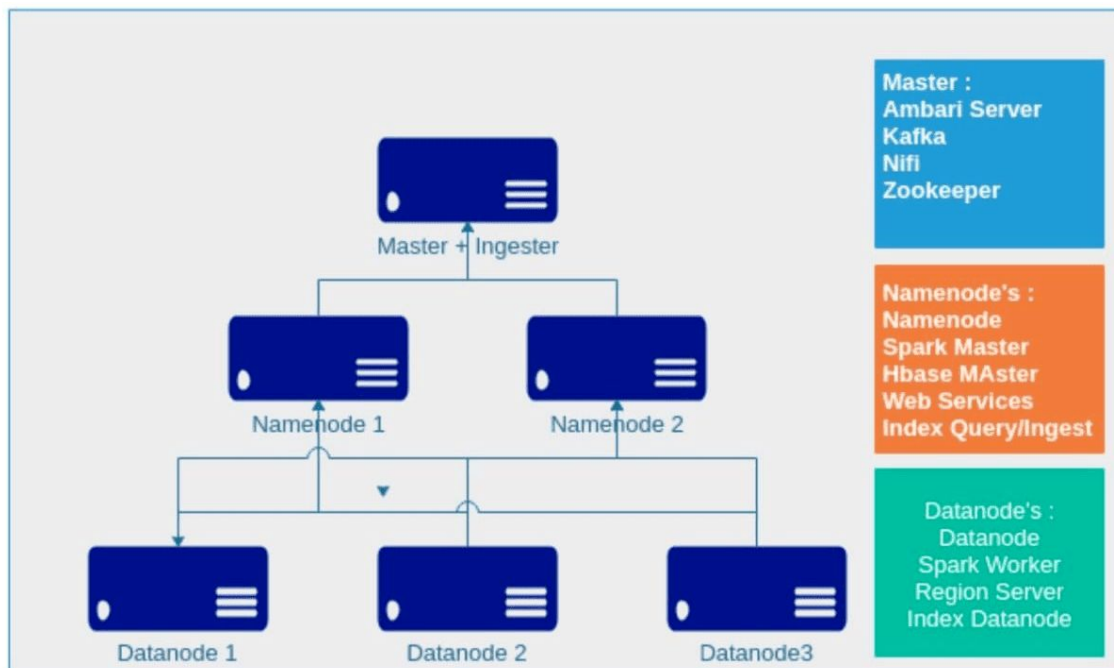
10. Spring

Spring framework merupakan kerangka kerja yang bersifat open-source. Spring digunakan sebagai framework web services. Spring mengimplementasikan beberapa teknologi IoC (Inversion of Control) dan DI (Dependency Injection) ke dalam sebuah MVC (Model View Controller). Spring juga menggunakan konsep AOP (Aspect

Oriented Programming) untuk memecah logika dari kode menjadi beberapa bagian (disebut concerns) untuk menjalankan pengembangan aplikasi yang terpadu.

11. Front-End

Data yang telah disiapkan web service selanjutnya dikirim ke Front-End. Stack Front-End yang digunakan yaitu Angular, React, Vue, dan Android. Pada sisi front-end data disajikan dalam bentuk grafis yang interaktif dan mudah digunakan.



1. MasterNode

Tugas Master adalah untuk mengkoordinasikan directory dan lokasi file block, memantau Namenode dan semua perubahan metadata. pada master node terdiri dari namenode dan juga *jobtracker*, *JobTracker* pada master node berfungsi untuk mengatur *job queue*. Jadi apabila ada sebuah aplikasi, maka aplikasi tersebut akan ditangani oleh JobTracker kemudian JobTracker mengatur salah satu bagian Slave yaitu *TaskTracker* yang mengerjakan masing-masing partisi data dari task yang diberi oleh aplikasi.

Misal ada aplikasi yang bisa menganalisis seberapa banyak kata “korupsi” muncul pada sosial media, lalu aplikasi akan mengontak Master, kemudian JobTracker yang sebenarnya adalah bagian dari MapReduce membagi job dan mendistribusikan ke Slave untuk diproses. Lalu Namenode akan memberitahukan indeks lokasi file kepada aplikasi.

Kelebihan lain dari cara kerja ini adalah ada salinan dari tiap data di masing-masing Slave/Node, atau saling tukar menukar salinan untuk menghindari apabila terjadi kerusakan pada Slave/Node. Sedangkan jika terjadi kerusakan justru ada di bagian Master, maka biasanya pada enterprise level disediakan 2 Master, yaitu Master Primary dan Master Backup untuk mengantisipasi terjadinya kerusakan pada Master Primary.

2. NameNode

NameNode dalam Arsitektur Apache Hadoop HDFS yang berfungsi untuk memelihara dan mengelola blok yang ada di DataNodes (node slave). NameNode adalah server yang sangat tersedia yang mengelola Namespace Sistem File dan mengontrol akses ke file oleh klien. Arsitektur HDFS dibuat sedemikian rupa sehingga data pengguna tidak pernah berada di NameNode. Data berada di DataNodes saja.

Jadi di dalam Namenode (salah satu bagian dari Master, yang sebenarnya adalah bagian dari HDFS) tidak berisi data, melainkan bagian ini hanya menyimpan indeks lokasi data yang kemudian data tersebut dipartisi menjadi kecil-kecil lalu disimpan di masing-masing Datanode.

3. Datanode

datanode adalah node-node dalam kluster yang bertindak sebagai slave, Datanode bertanggung-jawab menyimpan blok-blok data yang ditujukan kepadanya, dan secara berkala melaporkan kondisinya kepada namenode. Jadi, namenode seperti manager yang mengatur dan mengendalikan kluster.

Setiap data yang disimpan pada HDFS memiliki lebih dari satu salinan, yang disebut sebagai Replication Factor (RF). Secara default nilai RF adalah 3, yang berarti satu file tersimpan di 3 datanode berbeda sehingga jika salah satu datanode rusak, maka file dapat diperoleh dari datanode lain. Datanode mengirimkan sinyal setiap 3 detik yang disebut heartbeat kepada namenode untuk menunjukkan bahwa datanode tersebut masih aktif. Apabila dalam 10 menit namenode tidak menerima heartbeat dari datanode, maka datanode tersebut dianggap rusak atau tidak berfungsi sehingga setiap permintaan baca/tulis dialihkan ke node lain. Dengan heartbeat, maka namenode dapat mengetahui dan menguasai kondisi kluster secara keseluruhan. Sebagai respon atas heartbeat dari datanode, selanjutnya namenode akan mengirimkan perintah kepada datanode.