

## **BAB 2**

### **TEORI DAN TINJAUAN PUSTAKA**

Pada bab 2 akan dibahas tentang dasar teori dan tinjauan pustaka yang digunakan dalam pembuatan Proyek Akhir ini.

#### **2.1 Dasar Teori**

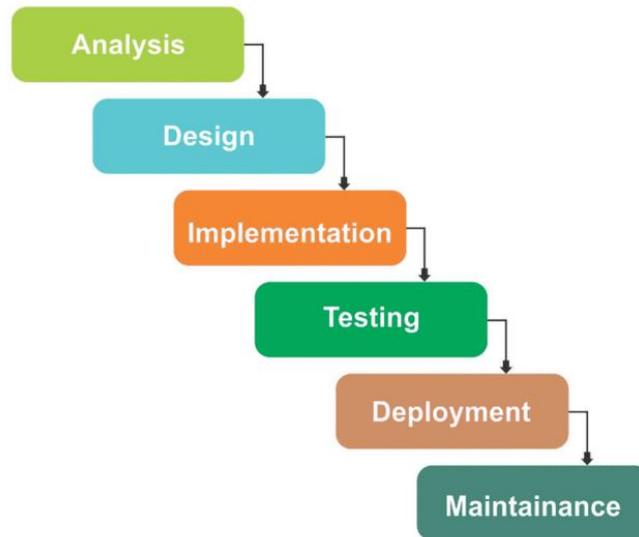
Dasar teori berisi tentang konsep atau pengetahuan, metode, dan tools yang digunakan untuk mendukung penyelesaian proyek akhir. Salah satu metode yang akan dibahas adalah metode Waterfall yang menjadi panduan utama dalam proses pengembangan Platform YAVA247. Selain itu, beberapa tools yang mendukung proses perancangan ulang agar lebih efisien juga akan dibahas.

##### **2.1.1 Platform Big Data Management**

Platform Big Data Management adalah suatu perangkat lunak atau sistem yang dirancang khusus untuk mengelola, mengolah, dan mengoptimalkan penggunaan data dalam skala besar serta dapat digunakan untuk memonitor kinerja perangkat secara real-time. Platform ini mencakup berbagai komponen dan layanan yang mendukung proses pengumpulan, penyimpanan, pengolahan, dan akses data.

##### **2.1.2 Metode Waterfall**

*Waterfall*, juga dikenal sebagai Software Development Life Cycle (SDLC), adalah salah satu metode pengembangan perangkat lunak yang mengikuti pola aliran, seperti air terjun. Metode *Waterfall* adalah pendekatan dalam pengembangan perangkat lunak dengan tahapan analisis kebutuhan, perancangan, implementasi, pengujian, dan pemeliharaan. Dalam metode ini, setiap tahapan pengembangan dilakukan secara berurutan, mengalir dari atas ke bawah.



Gambar 2.1 Tahapan Metode *Waterfall*

Tahapan Metode Waterfall dalam Pengembangan Platform YAVA247 terdiri dari beberapa tahap, antara lain:

- a) **Requirement Analysis** : Tahap ini dimulai dengan memahami kebutuhan dan tujuan dari perangkat lunak yang akan dikembangkan. Tim pengembang akan mempelajari kebutuhan dan persyaratan pengguna, serta menentukan fitur-fitur dan fungsi yang diperlukan
- b) **Design** : Spesifikasi kebutuhan dari tahap sebelumnya akan dipelajari dalam fase ini dan desain sistem disiapkan. Desain sistem membantu dalam menentukan perangkat keras (hardware) dan persyaratan sistem serta membantu dalam mendefinisikan arsitektur sistem secara keseluruhan. Pada tahap ini, dilakukan build source code services menggunakan Apache Maven sesuai dengan kebutuhan services YAVA247 versi Linux Ubuntu 22.04.
- c) **Implementation** : Tahap ini melanjutkan hasil desain source code service menjadi Debian Package. Proses ini melibatkan pengkodean debian control atau implementasi dan aktual dari perangkat lunak berdasarkan desain yang telah ditentukan sebelumnya. Tim developer menggunakan bahasa pemrograman dan alat pengembangan untuk menghasilkan perangkat lunak yang sesuai dengan spesifikasi desain.

d) **Testing dan Deployment** : Setelah file debian service selesai dibuat, tahap pengujian dilakukan untuk memastikan perangkat lunak berfungsi dengan baik. Hasilnya adalah perangkat lunak yang mampu memenuhi persyaratan pengguna. Tahap ini melibatkan proses implementasi aktual dari perangkat lunak berdasarkan desain yang telah ditentukan sebelumnya. Tim developer melakukan testing deployment pada node yang telah disediakan dan melakukan uji coba service yang telah di-deploy.

### 2.1.3 Apache Maven

Apache Maven adalah alat manajemen proyek yang digunakan dalam pengembangan perangkat lunak untuk mengotomatisasi proses pembangunan source code perangkat lunak. Maven menyediakan kerangka kerja yang konsisten dan standar untuk mengelola siklus hidup proyek perangkat lunak. Dengan Maven, pengembang dapat mendefinisikan struktur proyek, mengelola dependensi perangkat lunak, dan mengotomatisasi proses kompilasi, pengujian, dan paket distribusi. Maven juga mendukung integrasi dengan repositori Maven sentral, tempat penyimpanan dependensi perangkat lunak yang umumnya digunakan oleh komunitas pengembang.



Gambar 2.2 Logo Apache Maven

### 2.1.4 Apache Ambari

Apache Ambari merupakan sebuah framework yang dirancang khusus untuk memudahkan pengelolaan, konfigurasi, dan pemantauan cluster Apache Hadoop. Ambari menyediakan antarmuka pengguna (user interface) yang intuitif dan mudah digunakan, yang memungkinkan administrator untuk secara efisien mengelola kesehatan dan status cluster Hadoop. Selain itu, Ambari juga menyediakan fitur seperti pengaturan keamanan cluster, pemantauan kinerja, dan perangkat untuk analisis dan troubleshooting.

### 2.1.5 Apache Hadoop

Apache Hadoop adalah kerangka kerja perangkat lunak yang memungkinkan pemrosesan data terdistribusi berskala besar melalui kluster komputer yang sederhana. Komponen utamanya, seperti HDFS (Hadoop Distributed File System) dan YARN (Yet Another Resource Negotiator), membantu dalam manajemen sumber daya dan pemrosesan data. Hadoop sangat penting dalam ekosistem big data dan merupakan komponen utama dalam platform YAVA247.



Gambar 2.3 Logo Apache Hadoop

### 2.1.6 Apache Spark

Apache Spark merupakan teknologi komputasi kluster yang cepat, yang dirancang untuk perhitungan cepat. Hal ini didasarkan pada Hadoop MapReduce dan memperluas model dari MapReduce untuk efisiensi lebih banyak jenis perhitungan, yang mencakup query interaktif dan stream processing. Fitur utama Apache Spark adalah komputasi kluster di memory yang meningkatkan kecepatan pemrosesan aplikasi. Apache Spark dirancang untuk mencakup berbagai macam beban kerja seperti batch application, iterative algorithms, interactive queries dan streaming (Oliviandi, S., dkk 2018).



Gambar 2.4 Logo Apache Spark

### 2.1.7 Apache Hbase

Apache HBase adalah sistem penyimpanan big data sumber terbuka dengan pendekatan NoSQL dan terdistribusi. HBase merupakan basis data nonrelasional. Data dan permintaan didistribusikan di seluruh server dalam kluster HBase,

memungkinkan kueri hasil pada skala petabita data dalam waktu milidetik. HBase sangat efektif untuk menyimpan data nonrelasional, diakses melalui API HBase.



Gambar 2.5 Logo Apache Hbase

### 2.1.8 Apache Zookeeper

Apache ZooKeeper adalah layanan terpusat untuk memelihara informasi konfigurasi, penamaan, menyediakan sinkronisasi terdistribusi, dan menyediakan layanan grup. ZooKeeper juga menawarkan keandalan tinggi dengan replikasi data di beberapa server, memastikan ketersediaan layanan yang tinggi dan toleransi kesalahan. Dengan menyediakan fondasi yang kokoh untuk koordinasi dan sinkronisasi terdistribusi, ZooKeeper memainkan peran kunci dalam membangun aplikasi dan sistem yang dapat diandalkan dan tahan terhadap skenario yang kompleks dalam lingkungan terdistribusi.



Gambar 2.6 Logo Apache ZooKeeper

### 2.1.9 Git

Git adalah sistem kontrol versi terdistribusi yang digunakan untuk melacak perubahan dalam kode sumber selama pengembangan perangkat lunak. Dirancang untuk menangani proyek besar dengan kecepatan dan efisiensi, Git sangat berguna dalam kolaborasi tim, memastikan integritas kode, dan mengelola berbagai versi sebuah proyek. Git memungkinkan pengembang untuk bekerja secara lokal dan kemudian menyinkronkan perubahan mereka dengan server pusat. Ini memfasilitasi kolaborasi tim dalam proyek pengembangan perangkat lunak, seperti YAVA247.

## 2.2 Tinjauan Pustaka

Adapun beberapa penelitian sebelumnya yang dijadikan referensi untuk pembuatan Proyek Akhir ini antara lain :

Tabel 2.1 Tinjauan Pustaka

Judul Penelitian	Peneliti	Hasil
Deploying and Monitoring Hadoop, MapReduce Analytics on Single-chip Cloud Computer.	(Andreas-Lazaros Georgiadis, Sotirios Xydis, Dimitrios Soudris. 2016) (National Technical University of Athens)	Hasil penelitian menunjukkan bahwa SCC berhasil digunakan sebagai platform untuk menangani beban kerja Big Data dengan menggunakan infrastruktur yang terkonsolidasi dalam satu chip. Melalui proses pemantauan dan analisis performa, penelitian ini memberikan wawasan tentang efisiensi dan skalabilitas SCC dalam menjalankan aplikasi Hadoop dan MapReduce Analytics.
Ekosistem Apache Hadoop	Syauqi Ahsan, A., & Asmara, R. (2022).	Penelitian ini membuktikan penerapan analisis kinerja aplikasi big data dapat digunakan sebagai sarana optimasi desain infrastruktur big data.
EverAnalyzer, dan ekosistem Apache Hadoop.	Karamolegkos, P., Mavrogiorgou, A., Kiourtis, A., & Kyriazis, D. (2023).	Platform manajemen data EverAnalyzer menggunakan MapReduce dan Spark dalam memproses data sehingga user mendapatkan wawasan dari data tersebut untuk mengambil keputusan secara akurat.
Implementasi dan analisis hadoop element availability berdasarkan daemon log monitoring,	(Nurahdiatna Subagya, Adityas Wijajarto, Ahmad Almaari. 2021) (Universitas Telkom Bandung)	Hasil penelitian menunjukkan bahwa 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.