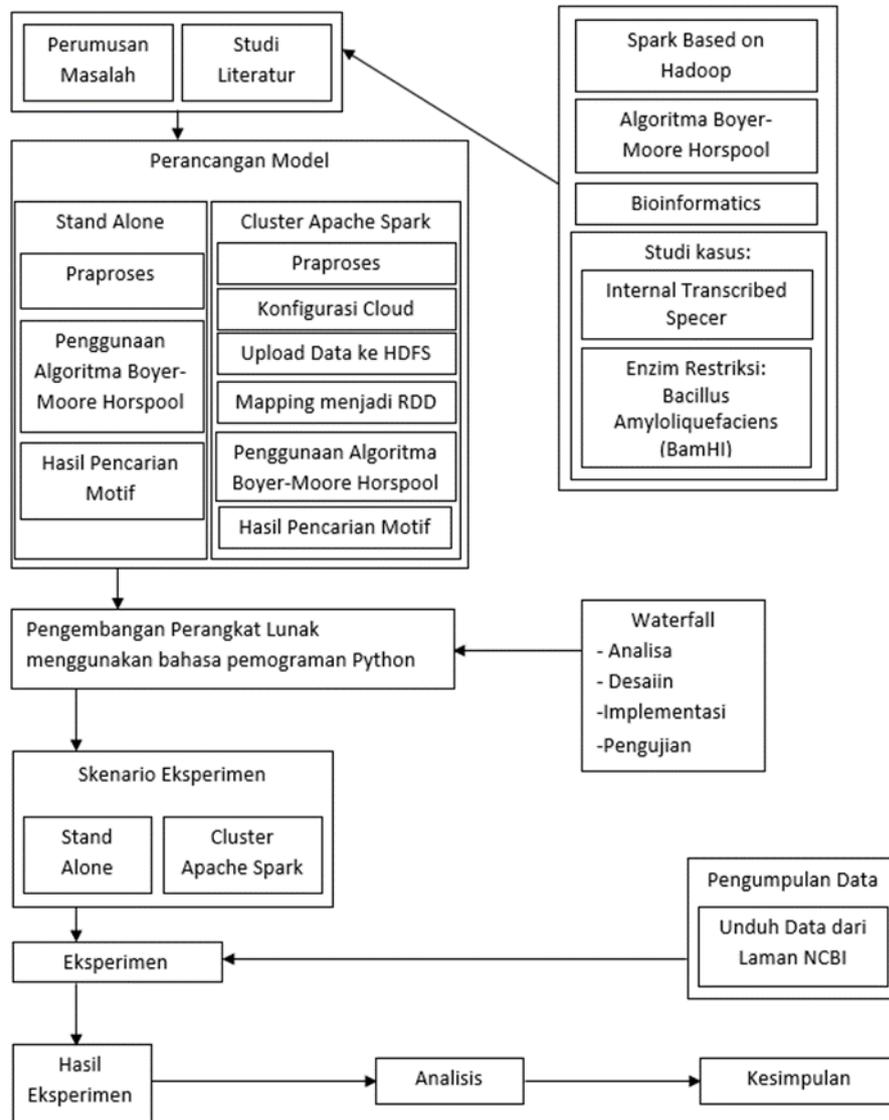


BAB III

METODOLOGI PENELITIAN

3.1. Desain Penelitian

Desain penelitian yaitu kerangka kerja yang digunakan untuk melakukan penelitian. Pada bagian ini penulis akan memaparkan kerangka kerja terkait penelitian dari memulai penelitian sampai dengan selesai. Desain penelitian digambarkan pada Gambar 3.1.



Gambar 3.1 Desain proses penelitian

1. Pada tahap pertama penulis melakukan diskusi untuk menemukan dan menentukan rumusan masalah yang dapat diangkat menjadi topik pada penelitian skripsi ini. Pada tahap ini pula penulis merumuskan tujuan dari penelitian yang berkaitan dengan rumusan masalah yang juga telah ditulis pada bab pertama skripsi ini.
2. Tahap selanjutnya penulis melakukan studi literatur terkait dengan topik yang telah ditentukan pada tahap pertama. Pada tahap ini dilakukan studi literatur tentang Apache Spark. Kemudian diperlukan juga studi literatur mengenai *string matching* dengan algoritma Boyer-Moore Horspool, pemrograman bahasa Python. Dan kemudian dilakukan studi literatur terkait dengan studi kasus yang akan digunakan yaitu *internal transcribed spacer* pada DNA tumbuhan, virus-virus pada DNA tanaman khususnya virus *Bacillus Amyloliquefaciens* (BamHI) yang menjadi fokus pada Biokomputasi penelitian.
3. Tahap berikutnya penulis melakukan pengumpulan data yang dapat diunduh dari *genbank* yang tersedia: FTP NCBI atau FTP Ensembl. Pada penelitian kali ini penulis mengunduh 62 file kromosom (kromosom random) berbagai jenis tanaman padi dari publikasi di laman FTP NCBI yang dapat diunduh pada <ftp://ftp.ncbi.nlm.nih.gov/genomes/>
4. Setelah itu penulis merancang model Apache Spark untuk pencarian motif pada sekuens menggunakan algoritma Boyer-Moore Horspool. Rancangan model yang dibuat dimulai dari praproses, penggunaan algoritma Boyer-Moore Horspool, menangkap hasil pencarian motif.
5. Selanjutnya dilakukan pembuatan program pencarian motif pada sekuens DNA dengan menggunakan algoritma Boyer-Moore Horspool yang diimplementasikan pada bahasa pemrograman Python dengan menggunakan Apache Spark dari model yang telah dirancang pada proses sebelumnya.
6. Setelah tahap pembuatan program selesai dilakukan analisis hingga pengujian, penulis membuat skenario eksperimen yang akan dilakukan selanjutnya.

7. Penulis melakukan eksperimen komputasi *stand alone* dan dengan Apache Spark dengan berbagai *node* untuk didapatkannya perbandingan dari setiap satu kali percobaan.
8. Langkah selanjutnya yaitu penulis melakukan analisis dari hasil eksperimen yang telah dilakukan pada tahap sebelumnya. Pada tahap ini penulis diharapkan mendapatkan grafik dari hasil eksperimen beserta dengan rekap indeks dari motif sekuens DNA dengan berbagai penggunaan jumlah *node* pada setiap percobaan.
9. Pada tahap terakhir, penulis akan mendokumentasikan hasil penelitian ini dan menyusun laporan penelitian dalam bentuk skripsi.

3.2. Alat dan Bahan Penelitian

Bagian ini menjelaskan secara detail alat dan bahan yang digunakan untuk melakukan penelitian.

3.2.1. Alat Penelitian

1. Perangkat Keras (*Hardware*) yaitu laptop dengan spesifikasi:
 - Processor Intel® Core™ i3-2330M CPU @ 2.20Ghz (4 CPUs), ~2.2GHz
 - Memory 6144MB RAM
 - Intel® HD Graphics 3000
 - HDD 730 GB
2. Perangkat Keras (*Hardware*) *cloud* untuk *cluster single node*, 3 *nodes*, 5 *nodes*, 11 *nodes*, 16 *nodes* dengan spesifikasi:
 - 4 vCPU
 - memori 15 GB
 - Storage / disk 32 GB
3. Perangkat lunak (*software*) sebagai berikut:
 - Sistem Operasi Linux 10 Pro 64-bit (10.0, Build 14393)
 - Python 3.7.1
 - Idle
 - Jupyter Notebook
 - Google SDK

- Microsoft Office Excel

3.2.2. Bahan Penelitian

Bahan yang diperlukan untuk melakukan penelitian yaitu data sekuens DNA tumbuhan khususnya tumbuhan padi yang dapat diunduh di Laman NCBI beralamat di <ftp://ftp.ncbi.nlm.nih.gov/genomes/> serta beberapa artikel dan jurnal pendukung sehingga dapat menunjang penelitian ini.

3.3. Metode Penelitian

Metode yang dilakukan dalam penelitian ini dibagi kedalam dua bagian, yaitu metode pengumpulan data dan metode pengembangan perangkat lunak.

3.3.1. Metode Pengumpulan Data

Penulis berusaha mendapatkan data yang valid dan mampu menunjang penelitian. Adapun metode pengumpulan data pada penelitian ini adalah sebagai berikut:

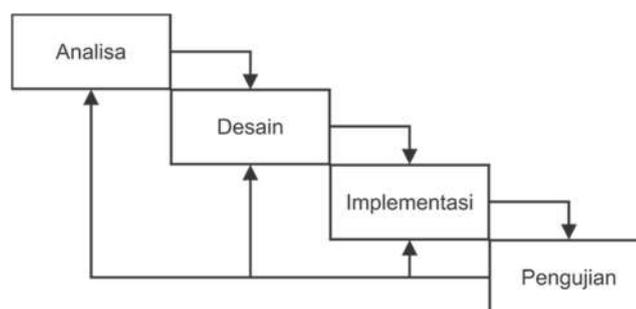
1. Studi Literatur

Studi literatur dilakukan dengan cara mempelajari teori dan konsep pendukung dalam penelitian yaitu Apache Spark, Hadoop, algoritma untuk *string matching* Boyer-Moore Horspool, dan yang terkait dengan studi kasus yaitu mengenai *internal transcribed spacer* dan virus-virus pada DNA tumbuhan khususnya virus *Bacillus Amyloliquefaciens* (BamHI) yang menjadi fokus penelitian Biokomputasi.

2. Mengunduh Data Sekuens

Pengumpulan data sekuens yang akan digunakan dalam penelitian ini diperoleh dari situs *genbank* yang terpercaya seperti NCBI. Data yang digunakan dapat diunduh dari sumber yang telah disebutkan pada sub bab bahan penelitian.

3.3.2. Metode Pengembangan Perangkat Lunak



Gambar 3.2 Model *Waterfall* (Sommerville, 2011)

Metode pengembangan perangkat lunak dilakukan dengan metode *waterfall*. Model SDLC air terjun (*waterfall*) sering juga disebut model sekuensial linear (*sequential linear*). Model *waterfall* menyediakan pendekatan alur hidup perangkat lunak secara sekuensial atau urut dimulai dari analisi, desain, pengkodean, pengujian dan tahap *support* (Sukamto & Shalahuddin, 2011). Pada penelitian ini, penulis menggunakan metode *waterfall* seperti pada Gambar 3.2 agar dapat dilakukan perbaikan terhadap kesalahan-kesalahan atau adanya kekuarangan di suatu tahapan. Berikut adalah tahapan pada model pengembangan perangkat lunak *waterfall* sesuai dengan gambar 3.2 menurut Ian Sommerville (2011):

1. *System Requirments Analysis and Definition* (Analisa)

Analisis adalah tahap menentukan aplikasi atau *software* seperti apakah yang akan dibuat. Analisis merupakan tahapan penetapan fitur, kendala dan tujuan sistem melalui konsultasi dengan pengguna sistem. Semua hal tersebut akan ditetapkan secara rinci dan berfungsi sebagai spesifikasi sistem. Analisis ini terdiri dari analisis kebutuhan dan analisis pembuatan sistem.

2. *System and Software Design* (Desain)

Dalam tahapan ini akan dibentuk suatu arsitektur sistem berdasarkan persyaratan yang telah ditetapkan. Dan juga mengidentifikasi dan menggambarkan abstraksi dasar sistem perangkat lunak dan hubungannya. Desain terdiri dari desain database, desain arsitektur sistem, dan desain antarmuka (*user interface*).

3. *Implementation and Unit Testing* (Implementasi)

Coding adalah tahap proses implementasi dari desain, dalam tahapan ini, hasil dari desain perangkat lunak akan direalisasikan sebagai satu set program atau unit program. Setiap unit akan diuji apakah sudah memenuhi spesifikasinya.

4. *Integration and System Testing* (Pengujian)

Proses *testing* atau pengujian dilakukan pada logika internal untuk memastikan semua pernyataan sudah diuji. Dalam tahapan ini, setiap unit program akan diintegrasikan satu sama lain dan diuji sebagai satu sistem yang utuh untuk memastikan sistem sudah memenuhi persyaratan yang ada. Setelah itu sistem akan dikirim ke pengguna sistem.