

PROJEK ILMIAH TAHAP AKHIR

Rampustakan, CINTA

**FAKULTI SAINS KOMPUTER DAN TEKNOLOGI MAKLUMAT
UNIVERSITI MALAYA**

TAJUK:

**PEMBERSIHAN DATA UNTUK GUDANG DAN PERLOMBONGAN DATA
(CLEANING DATA FOR WAREHOUSING AND MINING)**

CD4WAM

OLEH:

NAZATUSHIMA BT MOHD ARSHAD

WEK 990355

Laporan ini disediakan untuk Fakulti Sains Komputer dan Teknologi Maklumat, Universiti Malaya sebagai memenuhi sebahagian dari keperluan Ijazah Sarjana Muda Sains Komputer Dengan Kepujian sesi 2002/2003.

ABSTRAK

Pembersihan data bagi gudang dan perlombongan data atau nama ringkasnya *CD4WAM* merupakan satu kajian yang menggabungkan dua pangkalan data dari format berbeza menjadi satu format piawai. Pangkalan data yang telah digabungkan itu kemudiannya akan mengalami proses pembersihan data menggunakan teknik nilai yang hilang (missing values). Data yang telah dibersihkan itu kemudiannya akan dimasukkan ke dalam gudang data bagi memudahkan aktiviti perlombongan data. Proses pembersihan data merupakan satu proses mengenalpasti data-data yang bermasalah dan kemudiannya diperbetulkan menjadi data berguna (maklumat). *CD4WAM* ini akan bertindak selaku pembantu bagi pihak pentadbir gudang data dalam mengurus serta melancarkan tugas harian mereka. Selain itu, ia juga membantu pihak pengurusan atasan dalam membuat kesimpulan dan keputusan berasaskan kepada himpunan data yang banyak.

Terdapat 3 teknik yang digunakan dalam proses pembersihan data iaitu nilai yang hilang (missing values), kebisingan data (noisy data) dan ketidakkonsistenan data (inconsistent data). *CD4WAM* yang bakal dihasilkan ini akan memfokus kepada teknik nilai yang hilang. Bagi merealisasikannya, perisian-perisian seperti Visual Basic 6.0, Microsoft Access 2000 dan SQL Server 7.0 akan digunakan.

Selain itu *CD4WAM* dijangka dapat mengelakkan kesukaran capaian kembali maklumat yang diperlukan oleh pihak pentadbiran syarikat. Di samping

itu, ia juga memastikan agar hanya maklumat yang berguna sahaja yang terdapat di dalam gudang data.

Oleh itu adalah diharapkan agar *CD4WAM* ini mampu memberi manfaat kepada pihak pentadbir gudang data dan menjadi salah satu sumber penting bagi memperolehi maklumat serta penyelenggaraan data.

University of Malaya

PENGHARGAAN

Bersyukur saya ke hadrat Ilahi kerana dengan limpah kurnia-Nya saya dapat menyelesaikan Projek Ilmiah 2 ini dalam tempoh masa yang ditetapkan. Di sini saya ingin merakamkan jutaan penghargaan dan terima kasih kepada **Encik Teh Ying Wah** selaku penyelia yang banyak memberikan tunjuk ajar serta sokongan dalam membantu saya menyiapkan laporan Projek Ilmiah 2 ini. Ribuan terima kasih yang tidak tertinggi juga kepada **Puan Fariza Hanum Md Nasaruddin** selaku moderator di atas kerjasama yang baik serta cadangan-cadangan yang bernas bagi meningkatkan kualiti penghasilan **CD4WAM**.

Salam kasih selaut rindu buat kedua ayahbonda dan keluarga yang tercinta iaitu Mohd Arshad B Hj Ramli & Siti Sari Bt Hj Mokri serta Along sekeluarga. Tidak dilupa kepada rakan-rakan yang banyak membantu Ieda, Pojie, Nor, Noni, rakan-rakan serumah T135B dan rakan-rakan seperjuangan yang lain.

Akhir kata, ribuan terima kasih segunung impian diucapkan kepada semua individu-individu yang turut membantu saya secara langsung atau tidak langsung dalam menjayakan Projek Ilmiah 2 ini.

Sekian, terima kasih.

NAZATUSSHIMA BT MOHD ARSHAD

FAKULTI SAINS KOMPUTER DAN TEKNOLOGI MAKLUMAT

UNIVERSITI MALAYA

SESI 2002/2003

ISI KANDUNGAN

Halaman

Abstrak.....	i
Penghargaan.....	iii
Isi Kandungan.....	iv
Senarai Jadual.....	ix
Senarai Rajah.....	x

BAB 1 PENGENALAN

1.1 Pendahuluan.....	1
1.2 Rasional Sistem.....	2
1.3 Objektif Sistem.....	3
1.3.1 Objektif Akademik.....	3
1.4 Skop Sistem.....	4
1.5 Sasaran Pengguna.....	5
1.6 Ciri-Ciri Sistem.....	5
1.7 Perancangan Pembangunan Sistem.....	5
1.8 Hasil Akhir Yang Dijangkakan.....	10
1.9 Organisasi Bab.....	10

BAB 2 KAJIAN LITERASI

2.1	Pendahuluan.....	13
2.2	Gudang Data.....	14
2.2.1	Ciri-Ciri Bagi Gudang Data.....	16
2.2.2	Proses-Proses Dalam Gudang Data.....	17
2.2.3	Model Gudang Data.....	19
2.2.4	Kriteria Bagi Gudang Data.....	21
2.2.5	Komponen Dalam Persekitaran Gudang Data.....	24
2.2.6	Langkah-Langkah Dalam Gudang Data.....	25
2.3	Pembersihan Data.....	27
2.3.1	Klasifikasi Bagi Ralat Yang Sering Berlaku.....	28
2.3.2	Konflik Skema.....	29
2.3.3	Penggunaan Teknik Nilai Yang Hilang (Missing Values).....	32
2.4	Kajian Sistem Yang Sedia Ada.....	39
2.5	Rumusan.....	42

BAB 3 METODOLOGI SISTEM

3.1	Pendahuluan.....	43
3.2	Metodologi Pembangunan Sistem.....	43
3.2.1	Model Air Terjun Dan Prototaip.....	44
3.2.2	Pemilihan Model Air Terjun Dan Prototaip.....	46
3.2.3	Pemprototaipan.....	47
3.2.3.1	Masalah Berkaitan Pemprototaipan 'Evolutionary'.....	49
3.2.3.2	Pemprototaipan 'Throw-Away'.....	51
3.2.3.3	Isu-Isu Berkaitan Pemprototaipan 'Evolutionary'.....	52
3.3	Kitar Hayat Pembangunan Sistem (SDLC).....	53

3.4 Teknik Dan Kaedah Pengumpulan Maklumat	57
3.4.1 Perbincangan Dengan Penyelia Projek.....	57
3.4.2 Bahan Bacaan.....	58
3.4.3 Melayari Internet.....	58
3.5 Rumusan.....	59

BAB 4 ANALISIS KEPERLUAN SISTEM

4.1 Pendahuluan.....	60
4.2 Keperluan Fungsian.....	60
4.3 Keperluan Bukan Fungsian	62
4.4 Alatan Pembangunan.....	63
4.4.1 Keperluan Perkakasan.....	63
4.4.2 Keperluan Perisian.....	64
4.4.2.1 Microsoft Visual Basic 6.0.....	64
4.4.2.2 Pendekatan Pengaturcaraan Microsoft Visual Basic.....	67
4.4.3 Peralatan Implementasi Pangkalan Data.....	68
4.4.3.1 Sistem Pengurusan Pangkalan Data Perhubungan	68
4.4.3.2 Microsoft Access 2000.....	69
4.4.3.3 Microsoft SQL Server 6.0.....	70
4.4.3.4 Visual Foxpro 6.0.....	71
4.4.3.5 Oracle.....	72
4.4.3.6 Perbezaan Antara SQL Server Dengan Oracle.....	72
4.5 Rumusan.....	73

BAB 5 REKABENTUK SISTEM

5.1	Pendahuluan.....	74
5.2	Rekabentuk Struktur Sistem.....	75
5.2.1	Aliran Proses.....	76
5.3	Rekabentuk Antaramuka Pengguna.....	77
5.3.1	10 Heuristik Kebolegunaan.....	79
5.3.2	Perancangan Rekabentuk Antaramuka Sistem.....	81
5.4	Rumusan.....	84

BAB 6 IMPLEMENTASI DAN PERLAKSANAAN

6.1	Pendahuluan.....	85
6.2	Persekitaran Implementasi Dan Perlaksanaan.....	85
6.3	Faktor-Faktor Berkaitan.....	86
6.4	Pengaturcaraan <i>CD4WAM</i>	87
6.4.1	Pengaturcaraan Bermodul.....	87
6.4.2	Pengaturcaraan Berstruktur.....	88
6.5	Pendekatan Dalam Mengimplementasikan Pengaturcaraan.....	88
6.5.1	Pautan.....	89
6.5.2	Kebolehfahaman.....	89
6.5.3	Kebolehsuaian.....	90
6.6	Penggunaan Utiliti Dan Rutin Sepunya.....	90
6.7	Perlaksanaan Proses Pengaturcaraan.....	91
6.8	Hasil Fasa Implementasi Dan Perlaksanaan.....	92
6.9	Rumusan.....	93

BAB 7 PENGUJIAN DAN PENYELENGGARAAN

7.1	Pendahuluan.....	94
7.2	Peringkat Pengujian.....	94
7.2.1	Pengujian Unit.....	95
7.2.2	Pengujian Integrasi.....	95
7.2.3	Pengujian Sistem.....	95
7.3	Penyelenggaraan <i>CD4WAM</i>	99
7.4	Keperluan Penyelenggaraan.....	99
7.5	Rumusan.....	100

BAB 8 PENILAIAN SISTEM

8.1	Pendahuluan.....	101
8.2	Masalah Dan Penyelesaian.....	101
8.3	Kelebihan <i>CD4WAM</i>	105
8.4	Keterbatasan <i>CD4WAM</i>	106
8.5	Peningkatan <i>CD4WAM</i> Di Masa Hadapan.....	107
8.6	Cadangan.....	108
8.7	Pengajaran Dan Pengalaman.....	109
8.8	Kesimpulan.....	111

BIBLOGRAFI	113
-------------------------	-----

LAMPIRAN

i-MANUAL PENGGUNA

ii-KOD ATURCARA

SENARAI JADUAL

Bil	Jadual	Huraian
1	Jadual 1.1	Aktiviti Bagi Setiap Fasa Pembangunan
2	Jadual 1.2	Jadual Aktiviti Bagi Pembangunan Keseluruhan Projek
3	Jadual 2.1	Contoh Bagi Kehilangan Data Dan Ralat Dalam Jadual
4	Jadual 2.2	Contoh Jadual Yang Mengalami Kehilangan Data Dan Penyelesaiannya Menggunakan Fungsi Penganggaran Linear Antara Dua Nilai Yang Diketahui.
5	Jadual 2.3	Data Asli
6	Jadual 2.4	Contoh Jadual Yang Menggunakan Pembolehubah <i>Null</i> Bagi Menggantikan “-”
7	Jadual 2.5	Contoh Jadual Yang Menggunakan Pembolehubah <i>0</i> Bagi Menggantikan “-”

SENARAI RAJAH

Bil	Rajah	Huraian
1	Rajah 1.1	Perancangan Pengurusan Projek <i>CD4WAM</i>
2	Rajah 2.1	Model Gudang Data
3	Rajah 2.2	Skema Gudang Data
4	Rajah 3.1	Model Air Terjun Dengan Prototaip
5	Rajah 3.2	Proses Pemprototaipan
6	Rajah 3.3	Kaedah Pemprototaipan
7	Rajah 3.4	Proses Bagi Pemprototaipan 'Throw-Away'
8	Rajah 3.5	Tujuh Fasa Dalam Kitar Hayat Pembangunan Sistem
9	Rajah 5.1	Aliran Proses Bagi <i>CD4WAM</i>
10	Rajah 5.2	Rekabentuk Antaramuka <i>CD4WAM</i>
11	Rajah 5.3	Antaramuka 1
12	Rajah 5.4	Antaramuka 2
13	Rajah 5.5	Antaramuka 1 Yang Telah Dipilih
14	Rajah 5.6	Antaramuka 2 Yang Telah Dipilih

PENGENALAN

1.1 PENDAHULUAN

Ledakkan maklumat dalam era perkembangan teknologi siber telah menyebabkan longgokan data yang terlampau banyak dan sukar diurus. Oleh itu, pengurusan maklumat yang berkesan amat dititikberatkan di samping menjadi satu panduan bagi menyokong keputusan yang bakal dilakukan oleh pihak pengurusan. Ini kerana pengurusan maklumat yang baik akan memberikan suatu hasil keputusan yang serupa.

Dengan mempertimbangkan keperluan-keperluan ini, maka pembersihan data menggunakan teknik data yang hilang (missing values) atau diringkaskan kepada *CD4WAM* dibangunkan bagi mengatasi pertindanan antara data yang wujud dalam gudang data. Di samping itu pengurusan maklumat yang baik juga harus ditekankan bagi menghadapi segala permasalahan yang bakal timbul. Oleh yang demikian, data-data yang tidak berguna akan menyebabkan kekeliruan dalam pengurusan gudang data dan akan menyukarkan capaian oleh pihak pentadbir gudang data.

Kesimpulannya, laporan ini akan membincangkan berkenaan objektif, skop dan perkara-perkara yang berkait secara langsung atau tidak langsung dalam membangunkan *CD4WAM*.

1.2 RASIONAL SISTEM

Walaupun perkembangan teknologi maklumat samada dari aspek perkakasan mahupun perisian berkembang pesat, namun pengaplikasiannya masih tidak menyeluruh. Oleh itu *CD4WAM* dibangunkan bagi mengatasi kekangan-kekangan yang ada dan seterusnya mampu memenuhi ciri-ciri seperti berikut:

- 1) Menggunakan satu piawai bagi semua atribut yang terlibat bagi mengelakkan kekeliruan bagi set entiti yang terlibat
- 2) Dapat menjimatkan masa dari segi tempoh capaian dan menjana laporan akhir
- 3) Memudahkan tugas pihak pentadbir gudang data bagi aktiviti capaian kembali maklumat
- 4) Memudahkan proses pengemaskinian maklumat yang ada

1.3 OBJEKTIF SISTEM

Berikut merupakan senarai objektif-objektif yang telah ditetapkan:

- 1) Menggabungkan dua pangkalan data dari format yang berbeza menjadi satu format piawai koleksi bersepadu pangkalan data
- 2) Mengenalpasti ralat yang wujud dari sumber yang pelbagai
- 3) Menggunakan teknik nilai yang hilang (missing values) dalam proses pembersihan data
- 4) Mengubah data menjadi maklumat bagi memudahkan aktiviti perlombongan data
- 5) Membantu dan memudahkan pentadbir gudang data mengurus dan membuat capaian ke atas gudang data
- 6) Mengelakkan risiko kehilangan data

1.3.1 OBJEKTIF AKADEMIK

- 1) Menunjukkan keberkesanan penggunaan Microsoft Access 2000, SQL Server dan Visual Basic bagi merekabentuk *CD4WAM*
- 2) Mengaplikasikan konsep pengurusan projek dalam lingkungan dunia professional
- 3) Menguji kemahiran dalam konsep perancangan projek, penilaian serta pengujian

- 4) Memupuk satu sifat bertanggungjawab serta pendedahan terhadap dunia pembangunan sistem yang sebenar

1.4 SKOP SISTEM

Skop merupakan garis panduan bagi memastikan *CD4WAM* memenuhi keperluan projek. Ia berfungsi bagi melaksanakan fungsian berikut:

- 1) *CD4WAM* hanya boleh digunakan oleh pihak pentadbir gudang data bagi membantu mengurus perjalanan operasi organisasi
- 2) Menyediakan satu mekanisma bagi membuat capaian ke atas data organisasi yang telah tersedia
- 3) Menyediakan kualiti data yang konsisten supaya tidak menimbulkan kekeliruan dan keraguan
- 4) Mengubah data yang banyak menjadi maklumat berguna bagi membantu pengurusan sesebuah organisasi

Secara amnya, *CD4WAM* berupaya mengatur dan mengurus longgokan data yang banyak menjadi maklumat berguna bagi membantu pihak pentadbir gudang data mencapai maklumat organisasi.

1.5 SASARAN PENGGUNA

Sebagai langkah permulaan, sasaran pengguna *CD4WAM* hanyalah kepada pihak pentadbir gudang data sesebuah organisasi.

1.6 CIRI-CIRI SISTEM

Ciri-ciri *CD4WAM* yang digariskan adalah seperti berikut:

- 1) Kesediadaan
- 2) Kebolehpercayaan
- 3) Ketepatan
- 4) Kesempurnaan
- 5) Keselarasan

1.7 PERANCANGAN PEMBANGUNAN SISTEM

Proses pembangunan *CD4WAM* terbahagi kepada dua peringkat:

- 1) Peringkat awal (semester 1)

Peringkat awal ini bermula pada bulan Jun 2002 sehingga awal bulan September 2002. Peringkat ini terdiri daripada 2 fasa pembangunan iaitu:

- a) Fasa Analisis dan Keperluan Sistem
- b) Fasa Rekabentuk

2) Peringkat akhir (semester 2)

Setelah selesai proses-proses pada peringkat awal maka proses akhir pula diteruskan. Peringkat ini merupakan pelaksanaan sebenar *CD4WAM* yang mana telah ditakrifkan pada peringkat awal. Fasa-fasa yang terlibat pada peringkat ini adalah:

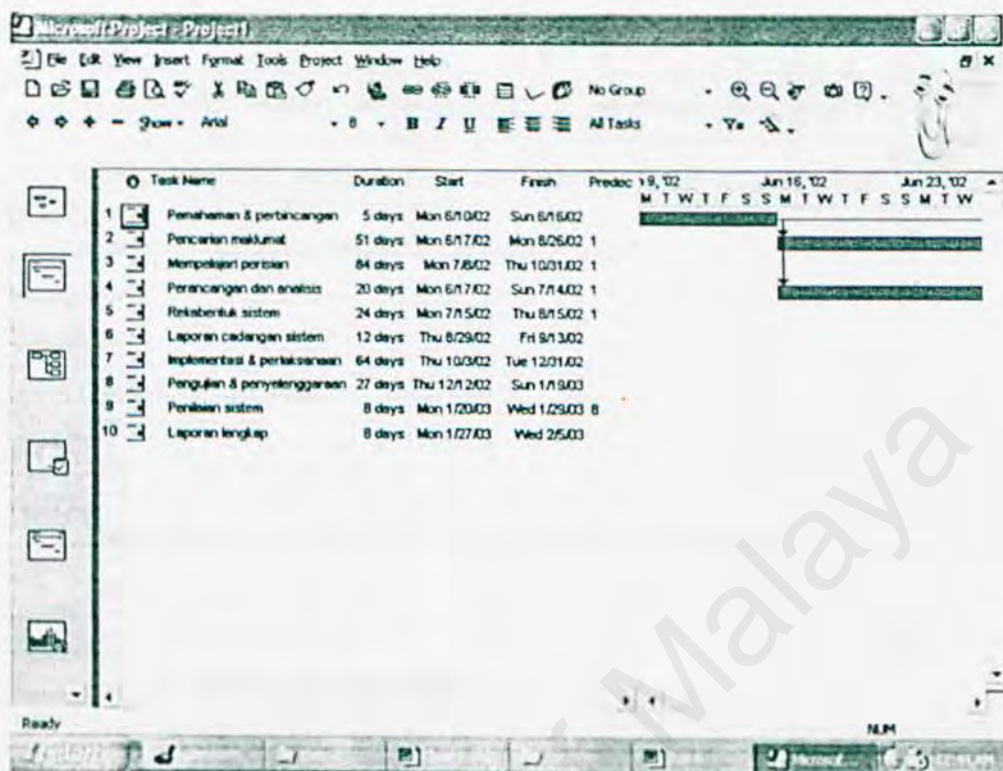
- c) Fasa Implementasi & Pelaksanaan
- d) Fasa Pengujian & Penyelenggaraan
- e) Fasa Penilaian Sistem
- f) Fasa Dokumentasi Akhir

BIL	FASA	AKTIVITI
1	Kajian awal dan analisis sistem	<ul style="list-style-type: none"> • Menentukan objektif dan skop • Proses pencarian maklumat • Menentukan keperluan sistem • Menyediakan diri dengan perisian yang bakal digunakan • Menyediakan perancangan projek • Memilih dan menentukan model pembangunan sistem
2	Rekabentuk sistem	<ul style="list-style-type: none"> • Merekabentuk pangkalan data • Merekabentuk antaramuka
3	Perlaksanaan	<ul style="list-style-type: none"> • Mempelajari perisian Visual Basic dan SQL Server • Melakukan pengkodan
4	Pengujian sistem	<ul style="list-style-type: none"> • Menguji modul-modul sistem
5	Penyelenggaraan sistem	<ul style="list-style-type: none"> • Melakukan perubahan terhadap sistem sekiranya terdapat permasalahan
6	Dokumentasi dan laporan	<ul style="list-style-type: none"> • Menyediakan laporan projek • Menyediakan manual pengguna

Jadual 1.1 Aktiviti Bagi Setiap Fasa Pembangunan

Fasa pembangunan	Jun 2002	Julai 2002	Ogos 2002	Sept 2002	Okt 2002	Nov 2002	Dis 2002	Jan 2003	Feb 2003
Pemahaman tajuk, perbincangan, pemerhatian (Fasa 1)	→								
Pencarian maklumat (Fasa 1)	→	→							
Mempelajari perisian yang bakal digunakan (Fasa 1)		→	→	→					
Perancangan awal serta analisis (Fasa 1)	→								
Laporan cadangan sistem (Fasa 1)			→	→					
Rekabentuk sistem (Fasa 1)					→	→	→	→	
Perlaksanaan pengkodan (Fasa 2)						→	→	→	
Pengujian dan penilaian sistem (Fasa 2)							→	→	
Penyelenggaraan sistem (Fasa 2)								→	→
Dokumentasi (Fasa 2)								→	→

Jadual 1.2 Jadual Aktiviti Bagi Pembangunan Keseluruhan
Projek



Rajah 1.1 Perancangan Pengurusan Projek CD4WAM

1.8 HASIL AKHIR YANG DIJANGKAKAN

Hasil penelitian terhadap proses pembangunan *CD4WAM* adalah dijangkakan bahawa ia akan memberikan hasil yang boleh membantu pihak pentadbir gudang data mendapat maklumat yang tepat dan jitu mengenai aktiviti organisasi mereka.

1.9 ORGANISASI BAB

Secara amnya, laporan ini terbahagi kepada 8 bab iaitu:

a) Bab 1 – Pengenalan

Bab ini memberikan gambaran dengan jelas tentang projek sebenar yang ingin dibangunkan. Ia menerangkan berkenaan dengan objektif, skop, sasaran serta perancangan pembangunan.

b) Bab 2 - Kajian Literasi

Bab ini membincangkan secara langsung mengenai gerak kerja awal dalam Bab 1. Kaedah-kaedah pencarian serta mendefinisikan kehendak maklumat dalam pelbagai persekitaran diaplikasikan dengan sebaik mungkin.

c) **Bab 3 - Metodologi Sistem**

Bab ini menerangkan tentang metodologi yang akan digunakan dalam proses pembangunan serta hasil analisis yang diperolehi

d) **Bab 4 – Analisis Keperluan Sistem**

Bab ini menerangkan mengenai alatan pembangunan yang akan digunakan serta keperluan-keperluan sistem seperti perkakasan dan perisian yang bakal digunakan.

e) **Bab 5 - Rekabentuk Sistem**

Bab ini memberi peluang kepada pembangun sistem menggunakan segala maklumat yang diperolehi untuk dipersembahkan dalam rekabentuk logikal sistem. Ia mempamerkan penekanan terhadap model-model pembangunan sistem. Ini kerana model tersebut merupakan garis panduan dalam pembangunan sistem.

f) **Bab 6 - Implementasi dan Perlaksanaan**

Bab ini menerangkan kaedah-kaedah implementasi yang digunakan dalam sistem dan pengaturcaraan. Di samping itu, aturcara-aturcara penting juga dinyatakan. Ini kerana aturcara yang dilampirkan merupakan aturcara sepunya.

g) **Bab 7 - Pengujian dan Penyelenggaraan**

Bab ini menerangkan mengenai kaedah-kaedah pengujian dari tahap pengujian unit, modul, integrasi serta sistem. Pengujian-pengujian ini penting bagi memastikan sistem boleh digunakan dengan baik tanpa ada kesalahan pada data yang dimasukkan. Tambahan pula, pengujian sistem yang baik dan teliti mampu memberikan suatu persembahan sistem yang berkualiti.

h) **Bab 8 – Penilaian Sistem**

Bab ini menerangkan secara keseluruhan mengenai masalah, cadangan dan kesimpulan sepanjang membangunkan sistem yang dibincangkan dalam laporan ini. Cadangan-cadangan yang dilampirkan di harap dapat memperbaiki situasi serta persekitaran pembangunan sistem ini bagi pelajar-pelajar akan datang.

KAJIAN LITERASI

2.1 PENDAHULUAN

Pembangunan sesebuah sistem yang konsisten dan boleh dipercayai memerlukan kajian yang terperinci dan juga perancangan yang rapi. Bagi memastikan pemahaman yang mendalam bagi sesuatu sistem yang ingin dibangunkan, kajian yang meluas dilaksanakan dalam pelbagai aspek melalui perbincangan, perpustakaan, internet dan bilik dokumen di Fakulti Sains Komputer dan Teknologi Maklumat

Apa yang lebih penting dalam kajian literasi ini adalah melengkapkan diri pembangun *CD4WAM* dengan pengetahuan berkaitan kekangan dan kelebihan beberapa alatan pembangunan. Ini adalah cabaran sebenar sebelum keputusan akhir dapat dicapai. Kajian literasi ini secara tidak langsung dapat membantu pembangun *CD4WAM* memilih alatan pembangunan yang bersesuaian bagi membangunkan *CD4WAM*.

Dalam proses pembersihan data menggunakan teknik nilai yang hilang, pelbagai kajian telah dilakukan bagi menambahkan lagi pemahaman mengenai teknik yang bakal digunakan.

2.2 GUDANG DATA

Kemunculan organisasi-organisasi pada masa kini dalam pelbagai bidang menghasilkan jumlah data yang besar untuk diuruskan. Sebagai contoh, organisasi yang mempunyai beratus atau beribu cawangan seperti dalam bidang insurans dan perbankan mungkin mempunyai data dari setiap cawangan mereka. Adalah wajar organisasi besar kini mempunyai struktur dalaman yang kompleks kerana data yang pelbagai tersebut adalah datang dari lokasi yang berbeza atau sistem pengendalian yang berbeza atau mungkin juga di bawah skema yang berbeza. Pentadbir pangkalan data memerlukan capaian dari pelbagai sumber. Dengan menyediakan pertanyaan bagi sumber individu adalah tidak wajar dan tidak efisien. Tambahan pula, sumber-sumber bagi data tersebut hanya akan menyimpan data semasa tetapi pentadbir gudang data masih memerlukan capaian bagi data-data yang lepas. Oleh itu gudang data menyediakan satu penyelesaian bagi masalah tersebut.

Gudang data adalah sebuah tempat menyimpan maklumat yang dikumpul dari pelbagai sumber, disimpan di bawah satu skema yang unik pada satu tempat tunggal. Apabila diperolehi, data tersebut akan disimpan pada satu jangka masa yang panjang serta membenarkan capaian bagi data masa lampau.

Gudang data merupakan satu pangkalan data yang besar yang direkabentuk bagi pemprosesan yang cepat bagi pertanyaan, perancangan dan peringkasan data dan biasanya digunakan oleh organisasi yang bersaiz besar. Ia juga boleh dikatakan sebagai koleksi bersepadu pangkalan data berorientasikan subjek yang direka bagi menyokong fungsi sistem sokongan keputusan di mana

setiap unit bagi data adalah tidak meruap dan relevan bagi detik masa tertentu[1]

Matlamat sebenar gudang data adalah :

- a) Menyediakan capaian ke data organisasi
 - pertanyaan interaktif, analisis dan persembahan data
 - mekanisma untuk capaian data yang tidak boleh dicapai
- b) Menyediakan kualiti data yang konsisten
 - tidak mengira pandangan bagi sesuatu jenis data
- c) Mengubah data menjadi maklumat
 - permodelan dan peramalan
 - analisis data demografi secara potongan-dan-dadu (slice-and-dice)

Potensi gudang data boleh ditonjolkan sekiranya data yang sesuai telah dikumpul dan disimpan dalam satu gudang data. Gudang data merupakan Sistem Pengurusan Pangkalan Data Perhubungan (RDBMS) yang direkabentuk secara khusus untuk memenuhi keperluan sistem pemrosesan transaksi. Gudang data juga merupakan teknik baru yang berkuasa untuk menghasilkan capaian bagi data operasian dan mengatasi ketidakkonsistenan antara format data lampau yang berbeza. Oleh itu mengintegrasikan data melalui keseluruhan perniagaan tanpa menghiraukan lokasi, format atau keperluan komunikasi boleh menggabungkan banyak maklumat tambahan. Ia merupakan hubungan logikal antara apa yang dilihat oleh aplikasi Sistem Maklumat Eksekutif (EIS) bagi

sokongan keputusan pengurus dan aktiviti operasi syarikat. Dengan kata lain, gudang data menyediakan data yang sudah dipindahkan dan diringkaskan maka menjadikan ia lebih sesuai dan efisien bagi aplikasi Sistem Maklumat Eksekutif (EIS) dan Sistem Sokongan Keputusan (DSS).

2.2.1 CIRI-CIRI BAGI GUDANG DATA

Terdapat 4 ciri yang boleh menerangkan berkenaan gudang data:

- a) Berorientasikan objek
 - data diurus merujuk kepada subjek. Sebagai contoh, syarikat insurans menggunakan gudang data bagi menguruskan data pelanggan, premium, tuntutan serta produk-produk yang berbeza. Data tersebut diurus mengikut subjek yang hanya mengandungi maklumat yang berkaitan bagi menyokong pemprosesan keputusan
- b) Bersepadu
 - apabila data berada dalam banyak aplikasi berbeza dalam persekitaran operasi, kebanyakannya adalah tidak konsisten. Sebagai contoh, dalam satu aplikasi, jantina mungkin dikod sebagai 'm' dan 'f' dan mungkin juga '0' dan '1'. Apabila data dipindahkan dari persekitaran operasi ke dalam gudang data maka ia dianggap sebagai satu kumpulan pengkodan yang konsisten. Contohnya, semua data jantina akan dipindahkan kepada 'm' dan 'f'
- c) Berasaskan masa

- gudang data mengandungi satu tempat bagi menyimpan data yang berusia 5 hingga 10 tahun atau lebih yang akan diguna bagi tujuan perbandingan, trend dan peramalan. Data-data ini tidak akan dikemaskini
- d) Tidak berubah
- data tidak akan dikemaskini atau diubah dalam sebarang cara apabila ia memasuki gudang data tetapi akan dimuat dan dicapai sahaja

2.2.2 PROSES-PROSES DALAM GUDANG DATA

Fasa pertama dalam gudang data adalah memisahkan maklumat operasi semasa bagi memelihara keselamatan dan integriti bagi misi kritikal aplikasi OLTP yang mana memberikan capaian menyeluruh terhadap asas bagi data. Keputusan bagi pangkalan data atau gudang data akan menggunakan 100 Gb atau mungkin Terabit bagi ruang cakera. Apa yang lebih penting adalah teknik yang efisien bagi penyimpanan dan capaian semula dalam jumlah yang besar bagi maklumat.

Gudang data membolehkan capaian semula data dari pelbagai jenis pangkalan data operasi. Data tersebut kemudiannya akan dipindahkan dan dihantar ke gudang data bagi model yang dipilih (atau definisi pemetaan). Pemindahan data dan proses pergerakan akan sentiasa dilaksanakan dan kemaskini terhadap data dalam gudang adalah diperlukan supaya terdapat satu bentuk automasi bagi mengurus dan melaksanakan fungsi tersebut. Maklumat yang menerangkan model tersebut dan definisi bagi elemen sumber data

dipanggil 'metadata'. Metadata bermaksud bagaimana pengguna mencari dan memahami data dalam gudang dan ia merupakan bahagian yang paling penting. Metadata sekurang-kurangnya mengandungi:

- a) Struktur bagi data
- b) Algoritma yang akan diguna bagi peringkasan
- c) Pemetaan dari persekitaran operasi kepada gudang data

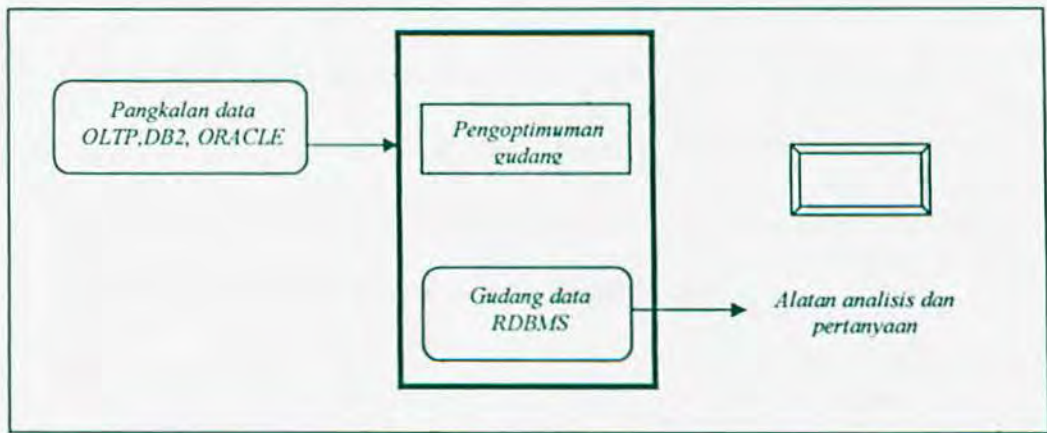
Pembersihan data adalah aspek yang sangat penting bagi menjana data yang efisien bagi gudang data di mana ia membuang aspek-aspek tertentu bagi data operasian seperti transaksi maklumat tahap rendah yang mana akan melambatkan semua jenis pertanyaan. Walaupun ia memerlukan maklumat tahap rendah, data sepatutnya dihasilkan dari sumber produksi pada sela masa tertentu tetapi proses pembersihan akan membuang salinan dan menyesuaikan perbezaan antara pelbagai bentuk koleksi data.

Apabila data telah dibersihkan ia kemudiannya dipindahkan ke gudang data yang mana merupakan satu pangkalan data yang sangat besar pada pelaksanaan yang tinggi samada Symmetric Multi-Processing (SMP) atau Massively Parallel Processing (MPP). 'Number crunching power' merupakan satu aspek lain yang penting bagi gudang data kerana kekompleksan yang terlibat dalam pemprosesan pertanyaan 'ad hoc' dan juga kerana kuantiti yang sangat luas bagi data yang diperlukan oleh organisasi untuk digunakan dalam gudang. Gudang data boleh digunakan dalam pelbagai cara yang berbeza. Sebagai contoh, ia boleh digunakan sebagai pusat simpanan di mana pertanyaan akan dilarikan atau ia boleh digunakan sebagai 'data mart'.

Data mart merupakan gudang yang kecil yang boleh dikekalkan bagi menyediakan subset bagi simpanan utama dan meringkaskan maklumat bergantung kepada keperluan bagi kumpulan atau jabatan tertentu. Pusat simpanan menggalakkan secara umum penggunaan struktur data yang ringkas dengan sedikit andaian tentang perhubungan antara pelbagai dimensi yang mana boleh meningkatkan pemprosesan pertanyaan dan ia juga boleh mempunyai struktur data. Kebanyakan vendor mempunyai produk yang menyediakan satu atau lebih fungsi gudang data yang diterangkan di atas. Walau bagaimanapun, ia boleh mengambil satu jumlah yang penting bagi kerja dan pengaturcaraan khusus untuk menyediakan kebolehsamaan yang diperlukan antara produk dari pelbagai vendor bagi membolehkan mereka untuk melaksanakan data yang diperlukan bagi proses gudang. Implementasi sering melibatkan satu campuran bagi produk dari pelbagai pembekal.

2.2.3 MODEL GUDANG DATA

Gudang data merupakan proses penghasilan dan perpindahan data operasi kepada data bermaklumat dan memuatkannya ke dalam satu pusat simpanan data atau gudang. Apabila data telah dimuatkan ia kemudian boleh dicapai melalui pertanyaan 'desktop' dan alatan analisis oleh pembuat keputusan. Model gudang data diilustrasikan seperti gambarajah di bawah:



Rajah 2.1 Model Gudang Data

Data perincian yang lama disimpan dalam beberapa bentuk bagi penyimpanan yang banyak. Ianya juga dicapai dan disimpan pada tahap perincian yang konsisten dengan data perincian semasa.

Kini peringkasan data merupakan data yang telah ditapis dari tahap terendah bagi perincian yang dijumpai pada tahap perincian semasa dan biasanya disimpan pada cakera penyimpanan. Apabila membina gudang data perlulah juga memberi perhatian bagi unit masa bagi menyelesaikan peringkasan dan juga kandungan atau atribut yang diperolehi oleh peringkasan data tadi.

Metadata merupakan komponen akhir bagi gudang data dan ia sememangnya dalam dimensi yang berbeza yang mana tidak sama seperti data yang dilihat dari persekitaran operasi tetapi ia diguna sebagai:

- a) Satu direktori bagi membantu penganalisis Sistem Sokongan Keputusan (DSS) memberitahu kandungan bagi gudang data

- b) Satu panduan bagi pemetaan data apabila data dipindahkan dari persekitaran operasi kepada persekitaran gudang data
- c) Satu panduan bagi algoritma yang akan digunakan bagi peringkasan antara data perincian semasa dengan data peringkasan terkini serta data peringkasan terkini dengan data peringkasan penting.

2.2.4 KRITERIA BAGI GUDANG DATA

Kriteria bagi gudang data RDBMS adalah seperti berikut:

- a) Pelaksanaan pemuatan
 - gudang data memerlukan pertambahan pemuatan bagi data baru pada masa asas tertentu tanpa masa tetingkap yang sempit (narrow time windows). Pelaksanaan bagi proses pemuatan sepatutnya diukur dalam 100,000 baris GB per jam dan seharusnya tidak menyekat jumlah data yang diperlukan oleh perniagaan
- b) Pemprosesan pemuatan
 - banyak langkah mesti diambil untuk pemuatan baru dan kemaskini data ke dalam gudang data termasuklah penukaran data, penapisan, memformat semula, pemeriksaan integriti, simpanan fizikal, pengindeksan dan mengemaskini metadata. Langkah ini mesti dilaksanakan secara berasingan
- c) Pengurusan kualiti data
 - pemindahan kepada pengurusan berasaskan fakta memerlukan kualiti data yang tinggi. Gudang mesti memastikan kekonsistenan

tempat, kekonsistenan global dan integriti rujukan. Walaupun sumber 'kotor' dan saiz pangkalan data yang besar semasa pemuatan dan persediaan bagi langkah yang sesuai ia adalah tidak memadai. 'Query throughput' adalah pengukuran bagi kejayaan aplikasi gudang data

d) Pelaksanaan pertanyaan

- pengurusan berasaskan fakta dan analisis 'ad hoc' tidak boleh lambat atau dihalang oleh pelaksanaan gudang data RDBMS, pertanyaan yang kompleks dan besar bagi operasi utama perniagaan mestilah lengkap dalam masa yang cepat

e) Kebolehskalaan Terabit (TB)

- saiz gudang data adalah meningkat pada tahap yang memberangsangkan. Pada masa kini, julatnya meningkat sehingga 100 GB dan TB - saiz gudang data adalah meningkat secara drastik. RDBMS sepatutnya tidak mempunyai sebarang keterbatasan. Ia seharusnya menyokong pengurusan selari dan modular. Ia mesti menyokong kebolehadan berterusan dalam kejadian pada titik kegagalan dan mesti menyediakan satu dasar mekanisma berbeza bagi mendapatkan semula. Akhir sekali, pelaksanaan pertanyaan mestilah tidak bergantung pada saiz pangkalan data tetapi lebih tepat pada kekompleksan bagi pertanyaan

f) Kebolehskalaan kumpulan pengguna

- capaian terhadap gudang data sepatutnya tidak dihadkan kepada kumpulan atasan. Pelayan RDBMS sepatutnya mampu menyokong

100 atau mungkin 1000 pengguna serentak apabila penyelenggaraan kebolehterimaan pelaksanaan pertanyaan

g) Rangkaian gudang data

- gudang data jarang wujud secara berasingan. Kebanyakan sistem gudang data beroperasi dalam satu rangkaian yang besar bagi gudang data. Pelayan sepatutnya menyediakan alatan yang mengkoordinasi pergerakan subset bagi data antara gudang. Pengurus gudang perlu mengurus dan mentadbir rangkaian bagi gudang dari satu lokasi fizikal

h) Pentadbiran gudang

- skala yang sangat besar dan 'time-cyclic nature' bagi gudang data memerlukan pentadbiran yang mudah dan fleksibel. RDBMS mesti menyediakan kawalan bagi mengimplementasi had bagi sumber, pengurusan perakaunan bagi mengenalpasti keperluan pengguna yang berbeza kelas dan aktiviti. RDBMS mesti menyediakan kawalan bagi pengesanan pemuatan kerja supaya sumber sistem akan dioptimumkan bagi pelaksanaan maksimum dan 'throughput'. Nilai yang mudah dilihat dan boleh diukur bagi implementasi sebuah gudang data adalah keterangan tanpa halangan iaitu capaian kreatif bagi data yang disediakan kepada pengguna akhir

i) Analisis Berdimensi Integrasi (Integrated Dimensional Analysis)

- kuasa bagi panduan pelbagai dimensi secara keseluruhannya diterima dan sokongan berdimensi mestilah semulajadi dalam gudang RDBMS bagi menyediakan pelaksanaan tertinggi bagi

alatan perhubungan OLAP. RDBMS mesti menyokong kepantasan iaitu penciptaan mudah bagi peringkasan biasa perkomputeran dalam gudang data yang besar. Ia juga sepatutnya menyediakan alatan penyelenggaraan untuk mengautomasikan penciptaan bagi pengagregatan perkomputeran. Perkiraan dinamik bagi pengagregatan sepatutnya konsisten dengan keperluan perlaksanaan interaktif

j) Fungsian Pertanyaan Maju (Advance Query Functionality)

- pengguna akhir memerlukan perkiraan analitikal yang maju, analisis berjujukan dan perbandingan serta capaian konsisten bagi memperinci dan meringkaskan data. Penggunaan SQL dalam alatan persekitaran pengguna/pelayan 'point-and-click' kadangkala tidak dapat dilaksanakan atau mustahil. RDBMS mesti menyediakan satu set yang lengkap bagi operasi analitikal termasuklah operasi berjujukan utama dan statistik

2.2.5 KOMPONEN DALAM PERSEKITARAN GUDANG DATA

a) Stor data operasian

- menstorkan data bagi satu set spesifik aplikasi operasian

b) Data mart

- subset bagi gudang data organisasi. Berperanan sebagai gudang data jabatan, wilayah dan fungsian

c) Metadata

- menyimpan maklumat mengenai data yang disimpan dalam gudang data

2.2.6 LANGKAH-LANGKAH DALAM GUDANG DATA

a) Perolehan

- data diperolehi dari pelbagai sumber

b) Pengintegrasian

- data yang dikumpul dalam langkah di atas akan digabung atau diintegrasikan kepada set data tunggal melalui kaedah integrasi skema

c) Penukaran

- set data yang telah diintegrasikan akan ditukar kepada format asas yang diperlukan bagi kegunaan pembersihan

d) Pembersihan

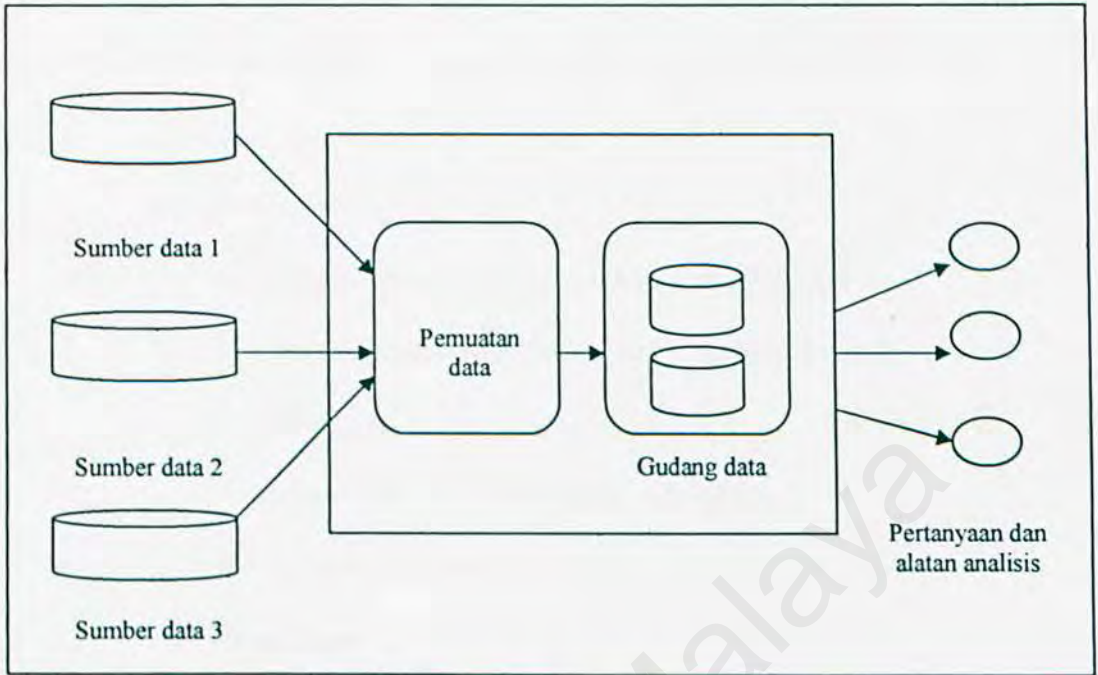
- pada peringkat ini, proses pembersihan dilakukan contohnya menghapuskan data yang bertindih serta mengesan dan membetulkan nilai data yang mempunyai ralat

e) Penyediaan

- selepas proses pembersihan, data kemudiannya disediakan dalam satu format yang diperlukan oleh gudang data

f) Pemuatan

- set data kemudiannya dimuatkan ke dalam gudang data dan sedia untuk digunakan



Rajah 2.2 Skema Gudang Data

2.3 PEMBERSIHAN DATA

Berikut adalah aplikasi-aplikasi perniagaan yang mana pembersihan data adalah penting:

- a) Pemasaran
 - mengenalpasti bentuk pembelian dari pengguna
 - mencari penyatuan di kalangan karektor demografik pelanggan
 - meramalkan tindakbalas melalui kempen
 - analisis 'market basket'
- b) Perbankan
 - mengesan bentuk penipuan penggunaan kad kredit
 - mengenalpasti pelanggan yang setia
 - meramalkan pelanggan yang gemar menukar kad kredit
 - menentukan jenis kad kredit yang disimpan oleh kumpulan pelanggan
 - mencari hubung kait yang tersembunyi di antara penunjuk kewangan yang berbeza
 - mengenalpasti peraturan stok pasaran melalui data pasaran lampau
- c) Insurans dan Kesihatan
 - analisis tuntutan
 - meramalkan pelanggan mana yang akan membeli polisi baru

- mengenalpasti bentuk gelagat bagi pelanggan berisiko
 - mengenalpasti gelagat penipuan
- d) Pengangkutan
- menentukan jadual pengagihan antara cawangan
 - analisis bentuk pemuatan

2.3.1 KLASIFIKASI BAGI RALAT YANG SERING BERLAKU

Sumber data yang diurus sendiri biasanya mempunyai skema yang berbeza. Secara nyata, ia mungkin menggunakan model data yang berbeza. Salah satu tugas penting dalam gudang data adalah melaksanakan integrasi skema dan menukarkan skema yang telah diintegrasikan sebelum ianya disimpan. Akhirnya, data yang disimpan dalam gudang data menjadi satu salinan bagi data dari pelbagai sumber.

Gudang data merupakan satu pangkalan data yang dapat melihat data secara terus dari sumber sistem. Kebanyakan pembangunan gudang data memerlukan masa yang agak panjang bagi proses perulangan yang melibatkan pengekstrakan, peralihan dan pemuatan data. Kewujudan ralat-ralat berikut memerlukan proses pembersihan:

- a) Data yang tidak lengkap
 - rekod atau medan yang hilang
 - rekod atau medan yang direkabentuk tetapi tidak direkodkan
- b) Tidak tepat
 - kesalahan kod

- kesalahan pengiraan
 - terdapat salinan rekod
 - salah memasukkan maklumat ke dalam sumber sistem
- c) Tidak dapat difahami
- kod yang tidak diketahui
 - fail pemprosesan perkataan
 - hubungan banyak-ke-banyak dan fail hierarki yang membenarkan pelbagai 'parents'
- d) Ketidakkonsistenan
- tidak konsisten bagi penggunaan kod yang berbeza
 - kod yang bertindih
 - kod yang berbeza dengan maksud yang sama
 - nama dan alamat yang tidak konsisten
 - pengiraan yang tidak konsisten
 - ketidakkonsistenan dari segi masa
 - ketidakkonsistenan penggunaan atribut

2.3.2 KONFLIK SKEMA

Konflik skema boleh wujud bagi perwakilan objek yang berbeza dalam skema yang berbeza dalam pertikaian cara. Dua jenis konflik yang wujud adalah konflik nama dan konflik struktur

1. Konflik nama

- objek yang mempunyai latar belakang aplikasi yang berbeza bagi organisasi yang sama merujuk kepada data yang sama menggunakan terminologi dan nama yang berbeza. Hasilnya akan memberikan ketidakkonsistenan yang munasabah dalam skema komponen. Masalahnya melibatkan dua jenis iaitu: Homonim (Homonyms) dan Sinonim (Synonyms).

Homonim berlaku apabila nama yang sama digunakan bagi dua konsep berbeza

Contoh: EQUIPMENT, dalam satu skema ia merujuk kepada komputer manakala yang lain merujuk kepada perabut (furniture)

Sinonim berlaku apabila konsep yang sama diterangkan melalui dua atau lebih nama. Contoh: CLIENT v/s CUSTOMER

Homonim boleh dikesan dengan membezakan konsep dengan nama yang sama dalam skema berbeza. Sinonim hanya boleh dikesan selepas spesifikasi luaran.

Satu jenis Homonim wujud bagi konsep yang sama yang mana sesuai pada nama tetapi tidak sesuai pada set yang serupa dengan contoh. Ianya boleh berlaku pada pelbagai tahap pengabstrakan.

2. Konflik struktur

- konflik ini wujud sebagai keputusan bagi pemilihan permodelan binaan atau kekangan integriti yang berbeza. Pengkelasannya adalah seperti berikut:

a) Konflik jenis

- wujud apabila konsep yang sama diwakilkan melalui binaan permodelan yang berbeza dalam skema berbeza. Sebagai contoh, satu kelas bagi objek diwakilkan sebagai entiti dalam satu skema dan sebagai atribut dalam skema yang lain

b) Konflik kelompok (dependency conflicts)

- wujud apabila satu kumpulan konsep adalah berkaitan di antara mereka dengan kelompok berbeza dalam skema berbeza.

c) Konflik kekunci

- kekunci berbeza ditentukan kepada konsep yang sama dalam skema berbeza. Sebagai contoh, SS# dan Emip mungkin kekunci bagi Pekerja dalam dua komponen skema

d) Konflik gelagat (behavioral conflicts)

- wujud apabila penyisipan berbeza bagi penghapusan polisi adalah disatukan dengan kelas yang sama bagi objek dalam satu skema yang nyata.

Data dalam dunia nyata adalah tidak lengkap, bising (noisy) dan tidak konsisten. Pembersihan data berusaha mengatasi nilai yang hilang (missing values), mengeluarkan kebisingan ketika proses mengenalpasti objek data yang tidak setara dengan model bagi data tersebut (outliers) dan memperbetulkan ketidakkonsistenan dalam data.

2.3.3 PENGGUNAAN TEKNIK NILAI YANG HILANG (MISSING VALUES)

Fokus utama bagi teknik ini adalah:

- a) Meminimakan data yang tidak bersih
 - fokus pada latihan, medan kerja dan masukan data
- b) Pertukaran dokumen
 - menyimpan fail program dengan penerangan bagi setiap perubahan yang dilakukan ke atas data mentah dan pembinaan pembolehubah
- c) Pengekalan data asli
 - memberi pilihan kepada pengguna bagi membuat pertukaran andaian
- d) Menggunakan teknologi pengiraan yang kukuh
 - menggunakan statistik berbanding pembolehubah yang tidak sensitif (contoh: *Num* adalah perwakilan bagi *Number*)

Bagi memenuhi nilai yang hilang bagi sesuatu atribut dalam jadual terdapat beberapa kaedah yang boleh digunakan iaitu:

1) Mengabaikan jadual tersebut

- ini hanya akan berlaku apabila label kelas didapati hilang (anggap sahaja tugas perlombongan melibatkan pengkelasan atau penerangan). Kaedah ini tidak efektif melainkan jadual tersebut mengandungi beberapa atribut dengan nilai yang hilang. Kaedah ini menjadi lebih buruk apabila peratusan bagi nilai yang hilang per atribut adalah banyak.

2) Menggantikan nilai yang hilang secara manual

- secara umum, kaedah ini berdasarkan masa dan tidak munasabah sekiranya diberi 1 set data yang besar dengan banyak nilai yang hilang

3) Menggunakan satu pembolehubah global

- menggantikan kesemua atribut nilai yang hilang dengan pembolehubah yang sama seperti satu label "*unknown*", "*not available*" atau "-". Jika nilai yang hilang digantikan dengan "*unknown*", maka program perlombongan akan menganggap ia melaksanakan

konsep yang berlainan sehinggalah ia mempunyai satu nilai yang biasa iaitu “*unknown*”.

4) Menggunakan atribut bermakna

- contohnya purata pendapatan bagi pelanggan Syarikat X adalah RM28,000 maka nilai ini akan digunakan bagi menggantikan nilai yang hilang bagi atribut dalam entiti *pendapatan*.

5) Menggunakan atribut bermakna bagi kesemua sampel yang merujuk kepada kelas yang sama seperti yang diberi pada jadual

- sebagai contoh, jika pengkelasan pelanggan merujuk kepada *risiko-kredit*, maka nilai yang hilang itu akan digantikan dengan nilai purata *pendapatan* bagi pelanggan dalam kategori risiko yang sama seperti jadual yang diberi

6) Menggunakan nilai kebarangkalian

- ia akan ditentukan dengan pengunduran, alatan berasaskan penyimpulan menggunakan satu formalisma Bayesian atau ransangan pokok keputusan. Sebagai contoh, menggunakan atribut pelanggan lain dalam set data atau membina satu pokok keputusan bagi meramalkan nilai yang hilang bagi *pendapatan*.

- kaedah yang sangat popular digunakan
- menggunakan kebanyakan maklumat dari data yang terdahulu bagi meramalkan data yang hilang
- dengan merujuk nilai bagi atribut yang lain dalam pengekalan nilai yang hilang bagi *pendapatan*, ia adalah cara yang baik bagi mengekalkan perhubungan antara *pendapatan* dengan atribut yang lain supaya terpelihara[2]

Pengenalpasti subjek	Nama	Umur	Fakta dunia sebenar	Pengkelasan
923C4	Ahmad	31	Ahmad berusia 31 tahun	<i>Benar</i>
782J7	Jenny		Jenny berusia 30 tahun	<i>Kehilangan nilai</i>
899P3	Peter	23	Peter berusia 21 tahun	<i>Ralat</i>

Jadual 2.1 Contoh Bagi Kehilangan Data Dan Ralat Dalam Jadual

Tarikh	Aset Semasa	Kredit Kerajaan
Feb-95	19452.6	837.4
Mac-95	18547.4	1175.2
Apr-95	18529.9	1252.3
Mei-95		
Jun-95		
Jul-95		
Ogo-95		
Sep-95		
Okt-95		
Nov-95		
Dis-95	17657.9	1577.9
Jan-96	17237.1	1142



Tarikh	Aset semasa	Kredit Kerajaan
Feb-95	19452.6	837.4
Mac-95	18547.4	1175.2
Apr-95	18529.9	1252.3
Mei-95	18420.8	1292.9
Jun-95	18311.7	1333.5
Jul-95	18202.6	1374.1
Ogo-95	18093.5	1414.7
Sep-95	17984.4	1455.3
Okt-95	17875.3	1495.9
Nov-95	17766.2	1536.5
Dis-95	17657.9	1577.9
Jan-96	17237.1	1142

Jadual 2.2 Contoh Jadual Yang Mengalami Kehilangan Data Dan Penyelesaiannya Menggunakan Fungsi Penganggaran Linear Antara Dua Nilai Yang Diketahui.

Tarikh	Jumlah Kutipan
Jan-96	45.20
Feb-96	0.01
Mac-96	15.54
Apr-96	-
Mei-96	-
Jun-96	0.09
Jul-96	0.08
Ogo-96	-
Sep-96	-
Okt-96	-
Nov-96	-
Dis-96	-

Jadual 2.3 Data Asli

Tarikh	Jumlah kutipan
Jan-96	45.20
Feb-96	0.01
Mac-96	15.54
Apr-96	null
Mei-96	null
Jun-96	0.09
Jul-96	0.08
Ogo-96	null
Sep-96	null
Okt-96	null
Nov-96	null
Dis-96	null

Jadual 2.4 Contoh Jadual Yang Menggunakan Pembolehubah *Null* Bagi
Menggantikan “-“

Tarikh	Jumlah kutipan
Jan-96	45.20
Feb-96	0.01
Mac-96	0.0
Apr-96	0.0
Mei-96	0.09
Jun-96	0.08
Jul-96	0.0
Ogo-96	0.0
Sep-96	0.0
Okt-96	0.0
Nov-96	0.0
Dis-96	0.0

Jadual 2.5 Contoh Jadual Yang Menggunakan Pembolehubah 0
 Bagi Menggantikan “_”

2.4 KAJIAN SISTEM YANG SEDIA ADA

Berikut adalah contoh-contoh laman web yang menawarkan sistem pembersihan data:

1. QM Soft [www.qmsoft.com]

Produk syarikat ini menggunakan teknik gabungan atau membersih berasaskan kepada:

- pelbagai kekunci
- kekunci lain
- 'soundex' (berasaskan prinsip 'phonetic')
- pendekatan papan kekunci
 - algoritmanya melihat kepada kedudukan karektor yang tidak sesuai dalam perhubungan bagi papan kekunci QWERTY
- pendekatan edit
 - algoritmanya menentukan berapa banyak karektor yang perlu diubah dalam nilai yang diberi bagi disesuaikan dengan nilai yang lain. Lebih sedikit karektor yang hendak ditukar, lebih tinggi kemungkinan penyesuaiannya
- algoritma frekuensi karektor
 - algoritmanya mengira bilangan kali setiap karektor digunakan dalam nilai yang diberi kemudian bandingkan ia dengan nilai kedua (sesuai bagi mencari pertukaran)

2. Harte-Hanks Data Technology [www.trilliumsoft.com]

Produk utama bagi Data Technology adalah 'Trillium Software System', satu koleksi kemudahan kerangka penulisan C bagi pengembalian data, analisis leksikal, 'data scrubbing' dan aktiviti penukaran data. Komponen bagi Trillium biasanya dilarikan pada MVS, Unix, Windows NT dan platform OS/2

- 'parser'
 - satu kemudahan bagi menanda, pempiawaian dan pengesahan data pelanggan
- 'matcher'
 - satu kemudahan bagi individu dan kesesuaian komersil dan pemilik rumah
- 'geocoder'
 - satu kemudahan yang mempertingkatkan alamat data dengan mengesahkan elemen alamat yang diketahui kepada direktori, mengesah atau menambah 9 digit poskod alamat, menyesuaikan komponen alamat kepada 'census coding directories' termasuklah poskod negeri, 'census tracts', kumpulan blok, 'Metropolitan Statistical Areas (MSAs), latitud dan longitud
- 'data browser'

- satu rekabentuk antaramuka pengguna (GUI) bagi menganalisis isu data integriti dalam satu suasana persekitaran kawalan kualiti

3. Innovative Systems Inc [www.innovgrp.com]

- 'innovative-dictionary'
 - satu sistem bagi menjana, mempiawai, mengesah dan mendokumenkan rekod pelanggan yang mana proses bagi rekod nama dan alamat pada satu bahasa tertentu, negeri tertentu, dan perkataan, frasa dan bentuk jadual tertentu
- 'innovative-match'
 - satu sistem bagi mengenalpasti rekod nama dan alamat yang terdapat salinan antara fail pelanggan
- 'innovative corp-match'
 - satu sistem bagi mengumpul rekod korporat berasaskan kesamaan dalam data nama dan alamat dan/atau medan pengguna tertentu seperti 'tax ID'
- 'innovative analyzer'
 - satu sistem yang melaksanakan pandangan keseluruhan bagi maklumat pelanggan dan masalah integriti data dalam satu pangkalan data maklumat pelanggan

2.5 RUMUSAN

Bagi mengatasi kekangan-kekangan yang terdapat pada manusia, proses-proses menyelesaikan masalah dianalisa sebagai sebuah sistem. Oleh itu, proses membuat keputusan menjadi lebih mudah dan hasil keputusan yang baik akan diperolehi.

CD4WAM yang bakal dibangunkan ini akan menitikberatkan konsep-konsep yang telah dibincangkan. Ini kerana longgokan data yang datang dari pelbagai sumber boleh mewujudkan pelbagai masalah kerana mempunyai perbezaan dari segi format, skema dan sebagainya. Oleh itu proses pembersihan data perlu mengambil kira semua aspek bagi memastikan data-data yang dihasilkan selepas proses pembersihan data adalah terjamin dan berkualiti.

METODOLOGI SISTEM

3.1 PENDAHULUAN

Fasa metodologi pembangunan sistem merupakan suatu fasa awal yang penting. Pada fasa ini, pembangun sistem perlu mengenalpasti tugas-tugas utama yang perlu dilakukan. Di antaranya adalah mengenalpasti keperluan fungsian dan juga keperluan bukan fungsian bagi membangunkan sesuatu sistem atau projek. Selain daripada itu, pembangun sistem juga perlu mempertimbangkan bahasa pengaturcaraan yang akan digunakan, senibina komputer, pangkalan data, perisian dan perkakasan yang bakal digunakan bagi membangunkan *CD4WAM*. Di samping itu juga, pendekatan-pendekatan yang telah dibincangkan dalam bab sebelum ini akan diketengahkan bagi memperlengkapkan keperluan sistem dan spesifikasi.

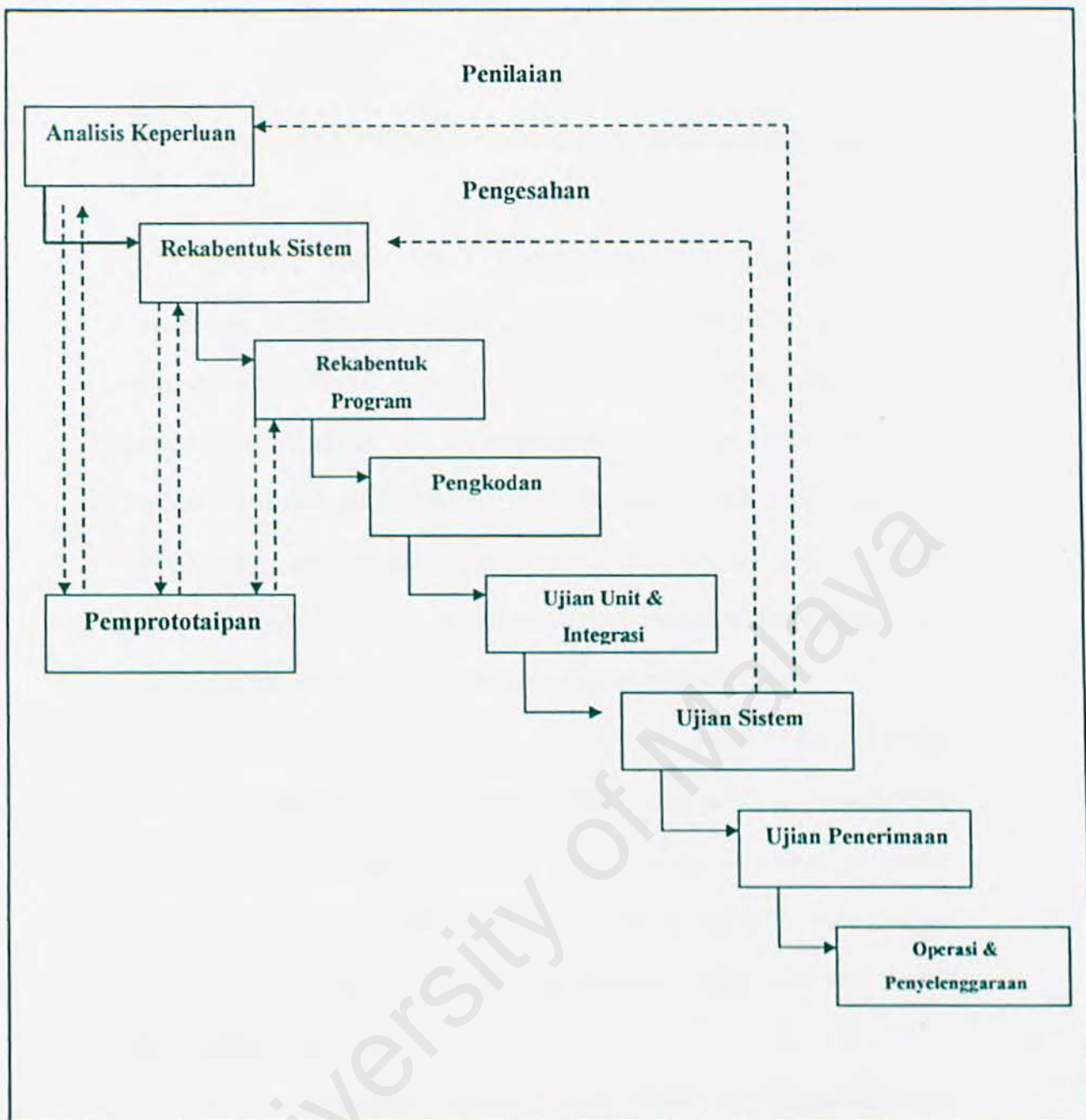
3.2 METODOLOGI PEMBANGUNAN SISTEM

Metodologi pembangunan sistem diperlukan sebagai garis panduan dalam usaha membangunkan sistem. Ia juga dikenali sebagai kitar hayat sistem di mana suatu set kaedah yang bermula dengan set keperluan dan menghasilkan sebuah sistem yang memenuhi kesemua keperluan yang diperlukan.

3.2.1 MODEL AIR TERJUN DENGAN PROTOTAIP

Oleh yang demikian, metodologi pembangunan yang dipilih adalah berasaskan Model Air Terjun dengan Prototaip. Pengaplikasian Model Prototaip dalam Model Air Terjun ini juga adalah untuk memperkenalkan mekanisme jaminan kualiti dalam proses pembangunan bahawa tiadanya penyimpangan daripada keperluan yang sepatutnya. Terdapat ciri-ciri yang mendorong dalam pemilihan model iaitu:

- a) Model ini popular dan digunakan oleh ramai para pembangun sistem
- b) Model ini berjjukan dan menunjukkan pemindahan proses dari suatu fasa ke fasa berikutnya
- c) Sekiranya terdapat sebarang perubahan atau kesilapan, pembangun sistem boleh kembali ke fasa sebelumnya tanpa menjejaskan pembangunan sistem
- d) Model ini jelas dan mudah difahami
- e) Pembangun sistem boleh menentukan secara kasar tempoh yang diperlukan bagi menyiapkan setiap fasa dan seterusnya jangka waktu untuk menyempurnakan keseluruhan projek dalam tempoh yang ditetapkan



Rajah 3.1 Model Air Terjun Dengan Prototaip

3.2.2 PEMILIHAN MODEL AIR TERJUN DENGAN PROTOTAIP

Kombinasi antara Model Prototaip dan Model Air Terjun akan memberikan penyelesaian terbaik bagi mengatasi masalah-masalah yang mungkin timbul kelak. Masalah bagi Model Air Terjun adalah proses pembangunannya di mana ia tidak membenarkan pengujian sehinggalah projek berakhir. Tambahan pula, aliran aktiviti bukanlah secara terus bagi sesuatu projek. Sistem penghantarannya kadangkala tidak berguna kerana ia tidak memenuhi keperluan sebenar pengguna. Ia tidak menyediakan satu usaha bagi menilai dan mencari langkah-langkah penyelesaian terhadap risiko.

Di sebalik perancangan, kos dan anggaran projek, prototaip adalah diluar skop pengalaman bagi ramai pengurus projek perisian. Prosedur bagi perubahan dan pengurusan konfigurasi mungkin tidak sesuai bagi perubahan pantas yang terdapat dalam prototaip. Pengurus mungkin menggunakan penekanan pada evolusi prototaip bagi mencapai kesimpulan yang segera tentang prototaip mereka.

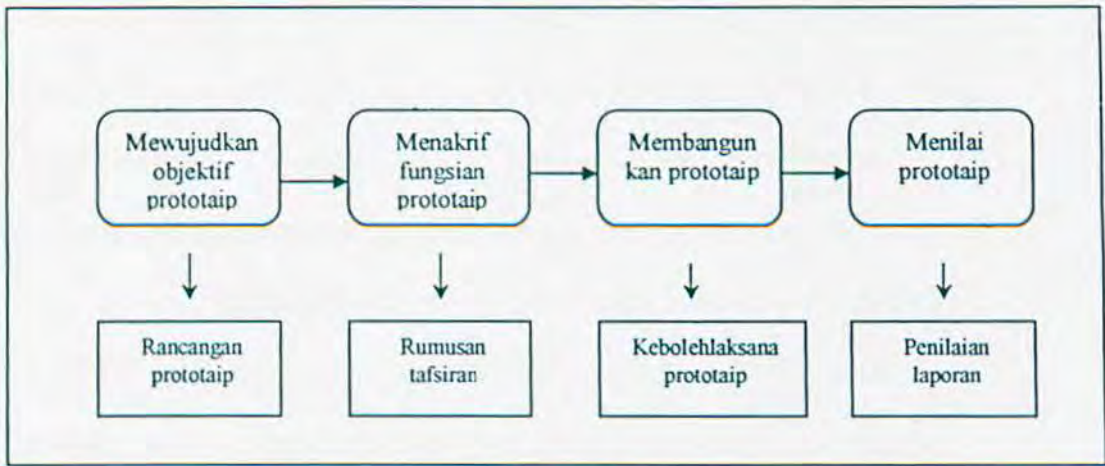
Bagi menyelesaikan masalah di antara Model Air Terjun dan Model Prototaip, maka Model Air Terjun dan Prototaip dipilih bagi pembangunan *CD4WAM*.

3.2.3 PEMPROTOTAIPAN

Pemprototaipan adalah versi awal bagi sesebuah sistem yang digunakan untuk menunjukkan konsep, mencuba pilihan-pilihan rekabentuk, memahami masalah dan mendapatkan penyelesaian-penyelesaian yang mungkin. Pemprototaipan dianggap sebagai aktiviti pengurangan risiko yang mana ia mengurangkan risiko keperluan.

Prototaip menyokong dua aktiviti kejuruteraan keperluan iaitu:

- a) 'Requirements Elicitation'
 - prototaip membolehkan pengguna mengeksperimen untuk melihat bagaimana sistem membantu tugas mereka. Pengguna boleh mendapatkan idea-idea baru bagi keperluan sistem.
- b) 'Requirements Validation'
 - prototaip dapat memperlihatkan ralat dan perkara-perkara yang tertinggal di dalam keperluan. Contohnya, fungsi yang telah dinyatakan di dalam spesifikasi keperluan apabila telah ditunjukkan melalui prototaip menunjukkan ianya kurang diperlukan.



Rajah 3.2 Proses Pemprototaipan

Terdapat 2 jenis pemprototaipan di dalam proses sistem iaitu:

a) Pemprototaipan 'Evolutionary'

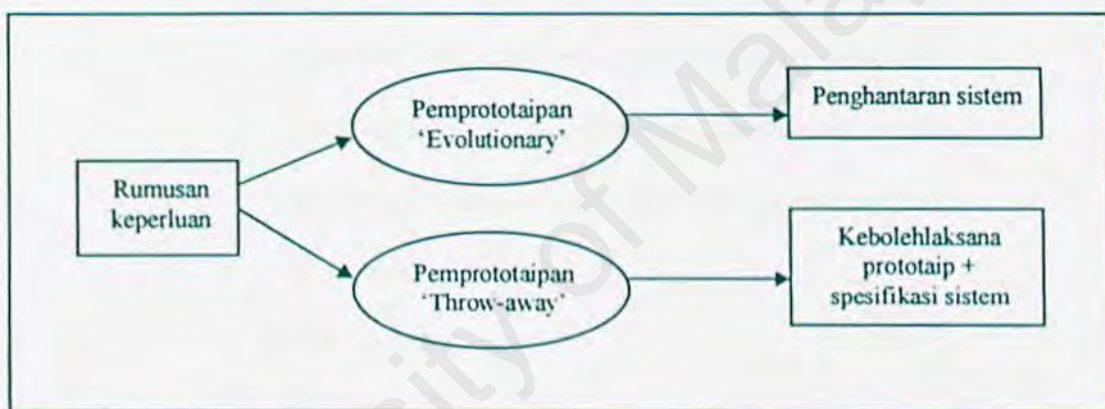
- adalah kaedah bagi sistem pembangunan yang mana prototaip awal dihasilkan dan diperbaiki melalui beberapa ulangan atau tahap sehingga sistem akhir dihasilkan

b) Pemprototaipan 'Throw-Away'

- prototaip bagi sesuatu sistem dibangunkan bagi membantu mengenalpasti masalah-masalah keperluan dan kemudiannya dibangunkan dengan menggunakan proses pembangunan yang lain

Bagi pembangunan *CD4WAM*, Pemprototaipan 'Throw-Away' akan digunakan sebagai Model Prototaip. Objektif bagi Pemprototaipan 'Throw-Away' adalah untuk mengesahkan atau menerbitkan keperluan-keperluan sistem. Pemprototaipan bermula dengan keperluan-keperluan yang tidak jelas

kerana keperluan ini perlu difahami. Keperluan yang jelas tidak diprototaipkan. Satu lagi perbezaan objektif di antara 2 jenis pemprototaipan ini adalah dari segi pengurusan kualiti bagi sistem tersebut. Pemprototaipan 'Throw-Away' mempunyai jangka hayat yang pendek. Prestasi dan kebolehpercayaan yang tidak baik boleh diterima asalkan prototaip tersebut dapat membantu memahami keperluan sistem. Tetapi bagi Pemprototaipan 'Evolutionary' yang mana seterusnya akan menjadi sistem akhir perlu dibangunkan dengan piawai kualiti yang sama seperti sistem-sistem lain di organisasi tersebut. Sistem perlulah mempunyai prestasi dan kebolehpercayaan yang baik dan sebagainya.



Rajah 3.3 Kaedah Pemprototaipan

3.2.3.1 MASALAH BERKAITAN PEMPROTOTAIPAN

'EVOLUTIONARY'

a) Masalah pengurusan

- prototaip berkembang dengan pantas, mungkin tidak kos-efektif untuk mengeluarkan sistem dokumentasi yang banyak. Ianya

bukan seperti model proses yang biasa . Contohnya, Model Air Terjun yang mana pengurus telah biasa dan mudah untuk mengawal kemajuan projek di sesuatu tahap. Pembangunan prototaip pantas memerlukan teknologi yang mungkin tidak biasa oleh staf yang sedia ada.

b) Masalah penyelenggaraan

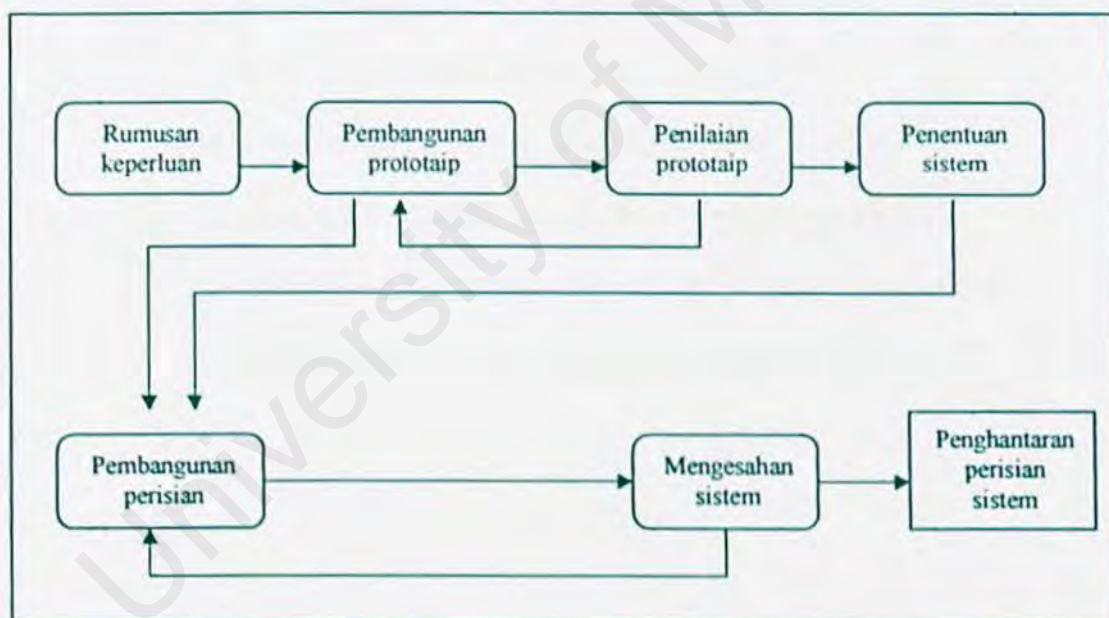
- perubahan yang kerap memberi akibat kepada struktur sistem prototaip tersebut. Selain dari pembangun asal mungkin adalah sukar bagi pembangun lain untuk mengubahnya. Lebih-lebih lagi sekiranya teknologi yang digunakan untuk menyokong pembangunan prototaip pantas sudah tidak digunakan lagi. Dengan ini, mungkin sukar untuk mencari orang yang dapat menyelenggara sistem ini

c) Masalah kontrak

- model kontrak yang biasa di antara pelanggan dan pembangun perisian adalah berdasarkan spesifikasi sistem. Apabila tiada spesifikasi, maka kemungkinan pelanggan kurang bersetuju untuk membayar kepada pembangun perisian mengikut masa yang diambil membangunkan sistem tersebut (fixed price contract).

3.2.3.2 PEMPROTOTAIPAN 'THROW-AWAY'

- a) Digunakan untuk mengurangkan risiko keperluan
- b) Prototaip dibangunkan dari spesifikasi awal untuk tujuan eksperimen dan kemudian tidak berguna lagi (discard)
- c) Prototaip 'Throw-Away' tidak dianggap sebagai satu sistem akhir kerana:
 - mungkin ada beberapa ciri sistem yang telah ditinggalkan
 - tiada spesifikasi bagi penyelenggaraan jangka panjang
 - sistem distruktur dengan cara yang tidak baik dan sukar untuk diselenggara



Rajah 3.4 Proses Bagi Pemprototaipan 'Throw-Away'

3.2.3.3 ISU-ISU BERKAITAN PEMPROTOTAIPAN

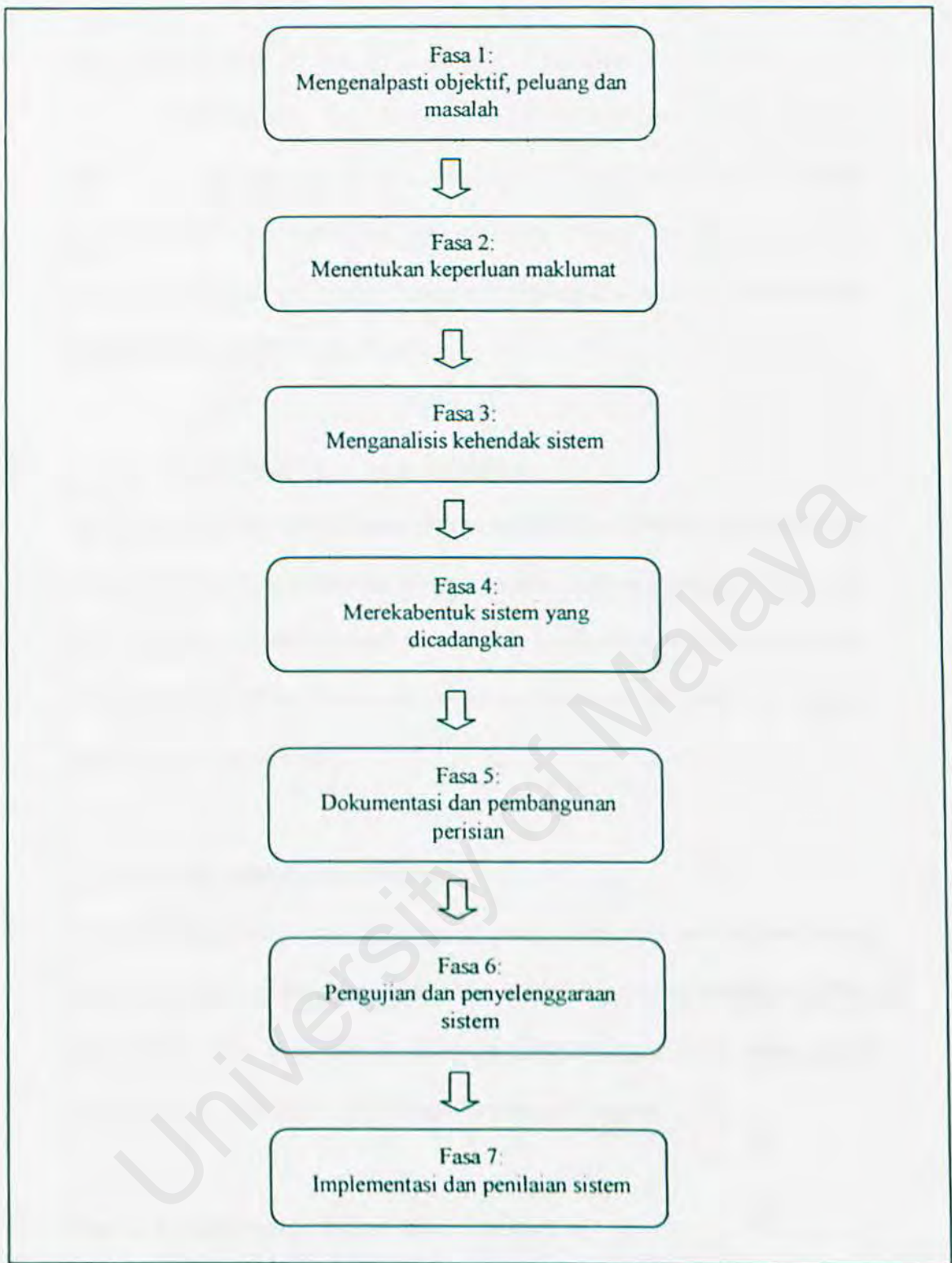
'EVOLUTIONARY'

- a) Pemprototaipan membolehkan sistem bersaiz kecil dan sederhana dibangunkan dan dihantarkan dengan pantas. Walau bagaimanapun bagi sistem yang bersaiz besar dan sekiranya juga melibatkan sub-kontraktor maka Pemprototaipan 'Evolutionary' adalah sukar dilaksanakan. Pemprototaipan 'Throw-Away' boleh dilaksanakan untuk memahami keperluan
- b) Proses pembangunan penokokan boleh menjadi alternatif kepada proses Pemprototaipan 'Evolutionary' bagi mengurangkan masalah-masalah yang timbul
- c) Proses pembangunan penokokan juga menyokong konsep prototaip dan pengurusan proses pembangunannya adalah dikendalikan mengikut cara yang biasa. Contohnya pelan dan dokumentasi dihasilkan bagi setiap penokokan [3]

3.3 KITAR HAYAT PEMBANGUNAN SISTEM (SDLC)

Kitar hayat pembangunan sistem (SDLC) merupakan satu fasa galakan bagi menganalisis dan merekabentuk keseluruhan sistem selain dari menggunakan kitar tertentu bagi analisis dan aktiviti pengguna. Terdapat tujuh fasa dalam SDLC iaitu:

University of Malaya



Rajah 3.5 Tujuh Fasa Dalam Kitar Hayat Pembangunan Sistem

Fasa 1: Mengenalpasti objektif, peluang dan masalah

Fasa ini adalah penting bagi kejayaan seluruh pembangunan sistem. Fasa ini perlu mengetahui apa yang berlaku dalam sistem. Kemudian, diikuti oleh punca utama masalah. Mengenalpasti objektif juga merupakan komponen yang penting bagi fasa ini. Sebagai permulaan, pembangun sistem mesti mengetahui apa yang cuba dihasilkan dari sistem.

Fasa 2: Menentukan keperluan maklumat

Fasa yang seterusnya adalah menentukan keperluan maklumat bagi penglibatan sesetengah pengguna. Beberapa alatan digunakan bagi mengenalpasti keperluan maklumat dalam sistem termasuklah persampelan dan menyelidiki data penting, temuramah, kaji selidik, pemerhatian gelagat pembuat keputusan dan suasana persekitaran serta prototaip

Fasa3: Menganalisis kehendak sistem

Terdapat alatan khas dan teknik yang dapat membantu pembangun sistem membuat penentuan keperluan. Ini termasuklah penggunaan gambarajah aliran data (DFD) yang mana terdiri daripada teknik struktur carta secara grafik meliputi input, proses dan output bagi fungsian perniagaan.

Fasa 4: Merekabentuk sistem yang dicadangkan

Dalam Fasa ini, pembangun sistem pada awalnya menggunakan pengumpulan maklumat bagi menyempurnakan rekabentuk logikal. Pembangun merekabentuk dengan tepat prosedur masukan data jadi sebarang data yang

dimasukkan adalah sentiasa benar. Pembangun juga merekabentuk sistem secara efektif menggunakan teknik bentuk kawalan dan rekabentuk skrin

Fasa 5: Dokumentasi dan pembangunan perisian

Sesetengah struktur teknik bagi dokumentasi dan rekabentuk perisian termasuklah kaedah HIPO, carta alir, carta Nassi-Schneiderman, gambarajah Warnier-Orr dan pseudokod. Semasa fasa ini, pembangun turut bekerjasama dengan pengguna bagi membangunkan dokumentasi bagi perisian termasuklah manual prosedur. Dokumentasi memberitahu pengguna bagaimana menggunakan perisian tersebut serta apa yang patut dilakukan sekiranya sebarang masalah wujud.

Fasa 6: Pengujian dan penyelenggaraan sistem

Sebelum sesuatu sistem tersebut boleh digunakan ia mesti diuji terlebih dahulu. Ini dapat menjimatkan kos pembangunan sekiranya masalah dapat dikenalpasti dengan lebih awal. Pengujian bagi punca masalah utama akan dijalankan secara bersiri dengan menggunakan sampel data dan data dari sistem semasa. Penyelenggaraan bagi sistem dan dokumentasinya hanya bermula dalam fasa ini.

Fasa 7: Implementasi dan penilaian sistem

Dalam fasa terakhir bagi pembangunan sistem, pembangun sistem akan membantu mengimplementasikan sistem yang ingin dibangunkan. Ini melibatkan latihan kepada pengguna bagi menggunakan sistem. Tambahan pula,

pembangun perlu merancang dengan lebih awal bagi melancarkan perubahan bagi sistem lama kepada sistem baru.

3.4 TEKNIK DAN KAEDAH PENGUMPULAN MAKLUMAT

Di dalam proses pengumpulan maklumat, pemahaman yang jelas terhadap sistem amat diperlukan bagi menghasilkan persembahan model yang baik. Terdapat pelbagai teknik yang telah digunakan bagi mengumpulkan segala maklumat yang berkaitan dengan pembangunan *CD4WAM*. Antara teknik yang digunakan adalah seperti mengadakan perbincangan dengan penyelia, membuat rujukan melalui bahan bacaan, melayari internet, melakukan kajian ke atas sistem yang sedia ada dan sebagainya.

3.4.1 PERBINCANGAN DENGAN PENYELIA PROJEK

Penyelia projek bertindak sebagai penasihat bagi melancarkan lagi *CD4WAM* yang akan dibangunkan. Perbincangan dengan penyelia adalah penting bagi mewujudkan persefahaman pendapat mengenai objektif, skop sistem dan juga keperluan perisian. Dengan wujudnya persefahaman dua hala ini, *CD4WAM* yang akan dihasilkan akan menjadi lebih bermutu dan mantap.

3.4.2 BAHAN BACAAN

Bahan bacaan tertentu dirujuk bagi memperolehi maklumat dan mendapatkan pemahaman yang jelas mengenai *CD4WAM* yang akan dibangunkan seperti membaca dan membuat perbandingan buku yang terdapat di perpustakaan, mengkaji buku-buku yang berkaitan dengan pangkalan data serta buku-buku mengenai perisian-perisian tertentu yang akan digunakan bagi membangunkan *CD4WAM*.

3.4.3 MELAYARI INTERNET

Melayari internet adalah merupakan kaedah atau cara yang lebih mudah, efisien dan efektif dalam membuat pencarian dan memperolehi sebarang maklumat khususnya berkenaan dengan *CD4WAM*. Ini termasuklah dari segi penggunaan perisian yang sesuai, maklumat mengenai pangkalan data, teknik pembersihan data dan sebagainya. Melalui kaedah ini juga, kajian dan analisis dijalankan bagi membuat perbandingan kesesuaian samada dua pangkalan data yang berbeza boleh digabungkan atau tidak.

3.5 RUMUSAN

Metodologi yang dipilih seharusnya sesuai dengan projek yang dibangunkan supaya ia tidak akan menimbulkan sebarang masalah ketika proses pembangunan sedang berlangsung. Setiap langkah yang terdapat dalam kitar hayat pembangunan sistem sepatutnya diikuti langkah demi langkah bagi memastikan proses pembangunan projek berjalan lancar dan mutu kerja lebih terjamin. Di samping itu pencarian maklumat haruslah berterusan samada secara perbincangan, bahan bacaan atau apa sahaja sumber maklumat bagi memantapkan lagi pemahaman dalam menghasilkan sebuah projek yang berkualiti.

ANALISIS KEPERLUAN SISTEM

4.1 PENDAHULUAN

Pencarian keperluan adalah merupakan suatu permulaan. Selepas keperluan dikenalpasti, seterusnya ia perlu dianalisis dan diklasifikasikan. Terdapat dua jenis keperluan yang perlu dianalisis iaitu keperluan fungsian dan keperluan bukan fungsian. Selain itu alatan pembangunan yang sesuai samada perisian atau perkakasan perlu di titik beratkan bagi memastikan proses pembangunan tidak menghadapi masalah di pertengahan jalan.

4.2 KEPERLUAN FUNGSIAN

Keperluan ini merujuk kepada fungsi-fungsi yang perlu ada di dalam sesuatu sistem supaya sistem itu dapat berfungsi sepenuhnya. Keperluan fungsian menerangkan suatu interaksi antara sistem dengan persekitaran. Dalam kes ini, ia merujuk kepada teknik nilai yang hilang dalam proses pembersihan data. Ini termasuklah dari segi penggunaan teknik pembersihan data, penyimpanan data dalam gudang data, perlombongan data, penggabungan pangkalan data dan sebagainya.

Keperluan fungsian bagi *CD4WAM* adalah seperti berikut:

- a) Memaparkan maklumat

- *CD4WAM* mampu memaparkan maklumat yang diinginkan oleh pengguna seperti senarai entiti bagi satu pangkalan data
- b) Meningkatkan kualiti data
 - Maklumat yang terdapat dalam gudang data adalah maklumat yang telah dibersihkan dan ini menjamin kualiti data
- c) Pencetakan
 - Membolehkan pencetakan maklumat yang perlu dicetak
- d) Autoriti
 - Hanya pengguna yang berautoriti sahaja dibenarkan untuk mencapai dan mengubah data
- e) Penukaran
 - membolehkan penukaran bagi format pangkalan data
- g) Ketepatan
 - pengguna dapat memperoleh data atau maklumat yang tepat dan tidak perlu diragui

4.3 KEPERLUAN BUKAN FUNGSIAN

Keperluan bukan fungsian merujuk kepada kekangan atau had-had terhadap fungsi yang ditawarkan oleh sistem. Ia juga mengambil kira had masa bagi proses pembangunan sistem.

Keperluan bukan fungsian bagi *CD4WAM* adalah seperti berikut:

a) Kebenaran

- *CD4WAM* yang akan dibangunkan mesti mengikut spesifikasi dan keperluan yang dinyatakan bagi mencapai objektif sistem

b) Kebolegunaan

- *CD4WAM* mudah untuk digunakan samada dari segi mengemaskini ataupun mencapai kembali maklumat

c) Keselamatan

- sumber data hendaklah disimpan dengan selamat memandangkan ianya melibatkan maklumat penting bagi sesebuah organisasi

d) Kebolehselenggaraan

- sistem boleh diselenggara sekiranya terdapat ralat yang tidak dikehendaki

4.4 ALATAN PEMBANGUNAN

Bagi menghasilkan *CD4WAM* yang menarik dan memenuhi setiap keperluan, aspek penting yang perlu dipertimbangkan adalah pemilihan perkakasan dan perisian yang sesuai. Pemilihan perlu dirancang dengan teliti supaya hasil gabungan bagi kedua-dua aspek tersebut akan dapat menghasilkan *CD4WAM* yang bermutu serta memberikan pulangan yang baik kepada pengguna yang menggunakannya.

4.4.1 KEPERLUAN PERKAKASAN

Perkakasan merupakan peralatan yang terdapat pada sesebuah komputer di mana ianya digunakan untuk membantu pembangun sistem dalam menghasilkan persembahan yang baik. Pembangun sistem akan memilih keperluan perkakasan yang sesuai di samping cuba untuk meminimumkan perkakasan tersebut.

Berikut adalah keperluan perkakasan yang diperlukan untuk pembangunan *CD4WAM*:

- a) Pentium 4 1.5Ghz
- b) Memori 256 SDRAM atau ke atas
- c) Sistem Pengoperasian Windows XP
- d) Cakera keras 20 GB dan ke atas
- e) Pemacu CD-ROM yang berkelajuan tinggi

- f) Monitor SVGA dengan sekurang-kurangnya 256 warna
- g) Tetikus dan papan kekunci

4.4.2 KEPERLUAN PERISIAN

Pembangunan *CD4WAM* akan dibangunkan dalam persekitaran sistem pengendalian Windows XP. Oleh itu, perisian-perisian yang akan digunakan mestilah yang serasi dengannya. Kajian telah dijalankan ke atas beberapa perisian tertentu bagi mengenalpasti perisian-perisian yang sesuai digunakan. Pemilihan perisian yang sesuai dan serasi dengan sistem pengendalian adalah penting kerana ia akan menentukan hasil akhir projek tersebut.

4.4.2.1 MICROSOFT VISUAL BASIC 6.0

Microsoft Visual Basic mencapai 90% kejayaan dalam perlaksanaannya. Ini menyebabkan ia sesuai digunakan sebagai platform untuk membangunkan sistem dengan lebih fleksibel dan berkesan. Ia juga mempunyai kebolegunaan yang tinggi dan pengaturcara sistem dapat memprogramkan sistem yang dikehendaki dengan agak mudah. Walaupun bahasa pengaturcaraan seperti C atau C++ mencapai peratusan yang tertinggi iaitu 95%, pembangunan sistem menggunakannya adalah sangat kompleks dan kurang sesuai.

Selain itu, ia juga merupakan gabungan di antara sebuah enjin yang berkuasa dengan persekitaran yang memudahkan kerja pengaturcaraan. Visual Basic juga dipilih kerana memiliki kriteria-kriteria yang masih dikekalkan

semenjak dari versi sebelumnya lagi di mana ia merupakan suatu keistimewaan bagi produk tersebut. Antara kriteria-kriteria tersebut adalah :

- a) Sistem pengaturcaraan Visual Basic membenarkan pereka merekabentuk aplikasi yang menarik dan berguna di mana ia menggunakan kaedah Antaramuka Pengguna Bergrafik (GUI) sepenuhnya. Ia akan menyebabkan proses rekabentuk skrin tidak mengambil masa yang panjang
- b) Wujudnya mod pengkompilan yang natif /asli. Dalam hal ini, ia mampu menjalankan proses pengkompilan secara keseluruhan berbanding *p-coded*. Mod pengkompil natif sesebuah aturcara mampu menampilkan hasil 20% lebih cepat berbanding pengkompilan secara *p-code* dalam Visual Basic
- c) Menyediakan kawalan ActiveX dengan menggunakan alatan pembangunan yang disediakan. Penciptaan kawalan-kawalan ini akan menghasilkan suatu fungsi kawalan yang lebih kurang sama jika ia dicipta dengan menggunakan Visual C++. Perbezaannya ialah pengaturcara boleh melakukan penciptaan tersebut dengan lebih cepat jika menggunakan Visual Basic
- d) Menyokong Objek Data ActiveX (ADO) model capaian data, alatan-alatan untuk manipulasi dan mengatur sambungan dan arahan-arahan serta menambahkan kawalan sempadan data di mana ia mengurus dan mengeluarkan sumber data tanpa memerlukan struktur pengaturcaraan yang kompleks

- e) Boleh diintegrasikan dengan kebanyakan format pangkalan data seperti Microsoft Access, Dbase, Visual FoxPro dan Paradox
- f) Serasi dengan sebarang perisian Windows
- g) Sesuatu aplikasi yang lengkap boleh direkabentuk dengan mengeksploitasikan sifat kekunci Windows termasuklah Antaramuka Pelbagai Dokumen (MDI), Pertukaran Data Dinamik (ODE), Grafik, Object Linking & Embedding (OLE) dan sebagainya
- h) Visual Basic boleh membentuk fail boleh laksana (fail EXE) dengan mudah menggunakan masa larian yang mampu dicapai dengan mudah
- i) Visual Basic membenarkan pengaturcaraan menulis, mengompil, melarikan dan menyahsilap (debugging) aplikasi di dalam persekitaran Windows tanpa perlu keluar ke DOS. Walaupun ciri ini ada pada perisian lain seperti Turbo Pascal untuk Windows tetapi ianya tidak secepat Visual Basic

Selain itu Visual Basic telah mengukuhkan lagi keistimewaannya dengan menambahkan kriteria-kriteria baru seperti berikut:

- a) Penghimpun kod semulajadi (Native Code Compiler)
 - Visual Basic mempunyai kebolehan untuk mengompil satu program kepada kod semulajadi seperti mana C++. Ini membolehkannya menghasilkan satu program yang lebih pantas. Walau bagaimanapun fail perpustakaan masa larian (runtime

library file) masih perlu untuk menyediakan satu program yang berfungsi sepenuhnya

b) Ciri-ciri pangkalan data yang baru

- Pengurus Data Visual (Visual Data Manager) merupakan ciri-ciri baru bagi Visual Basic. Ia memudahkan penyelenggaraan struktur pangkalan data seperti mana memasukkan dan penyuntingan data sebenar. Tambahan pula, Pengurus Data Visual membantu untuk mencipta, menguji dan menyimpan pernyataan SQL di dalam satu program

4.4.2.2 PENDEKATAN PENGATURCARAAN MICROSOFT VISUAL BASIC

Visual Basic menyediakan kemudahan untuk merekabentuk sistem aplikasi yang berstruktur dan bermodul. Di fasa pengkodan pula kod-kod akan ditulis bagi modul-modul berdasarkan fungsi-fungsi yang dilakukannya.

Pengaturcaraan bermodul adalah satu kaedah yang membahagikan sesuatu masalah yang kompleks kepada bahagian-bahagian kecil yang mudah di aturcarakan. Kebaikan menggunakan kaedah ini adalah kebolehan setiap modul aturcara dikod, dikompilasi, dilari dan diuji secara berasingan.

Bagi sistem ini, setelah pengkodan selesai dilakukan, kod akan terus ditulis, dikompilasi, dilari dan diuji bagi menentukan sesuatu modul itu benar-benar berfungsi. Kemudian barulah kod bagi modul lain ditulis.

4.4.3 PERALATAN IMPLEMENTASI PANGKALAN DATA

Analisis telah dilakukan bagi menentukan Sistem Pengurusan Pangkalan Data (DBMS) yang paling sesuai bagi menyimpan dan menguruskan koleksi data. Bagi memilih pangkalan data yang bersesuaian, pangkalan data tersebut mestilah memastikan keselamatan data yang disimpan. Pangkalan data merupakan aplikasi utama bagi kesemua perniagaan yang kritikal. Pemilihan pangkalan data yang tidak sesuai akan memberikan kesan yang teruk terhadap sesebuah organisasi.

4.4.3.1 SISTEM PENGURUSAN PANGKALAN DATA PERHUBUNGAN (RDBMS)

Sistem pengurusan pangkalan data perhubungan (RDBMS) merupakan satu piawai bagi bahasa pertanyaan. Idea bagi RDBMS dikemukakan dalam tahun 1970 dengan penghasilan model perhubungan oleh E.F Codd "Satu pangkalan data perhubungan membenarkan definasi bagi struktur data, penyimpanan, operasi capaian dan kekangan integriti. Dalam sesuatu pangkalan data, data dan perhubungan antara mereka adalah diuruskan dalam jadual. Jadual adalah satu koleksi bagi rekod dan setiap rekod dalam jadual mengandungi medan yang sama. Sesetengah medan mungkin direkabentuk sebagai kunci yang mana pencarian nilai tertentu bagi medan tersebut akan menggunakan pengindeksan bagi mempercepatkan pencarian. Rekod dalam

jadual berbeza mungkin dihubungkan jika ianya mempunyai nilai yang sama dalam satu medan tertentu dalam setiap jadual”[4]

4.4.3.2 MICROSOFT ACCESS 2000

Microsoft Access merupakan sistem pangkalan data perhubungan yang diguna bagi pangkalan data perhubungan. Seperti juga produk lain dalam kategori ini, Microsoft Access mampu menyimpan dan mencapai semula data, maklumat masa lampau dan mengautomasikan tugas berulang (seperti pengekalan akaun, pelaksanaan dan penjadualan kawalan inventori).

Microsoft Access juga merupakan aplikasi tettingkap yang berkuasa kemungkinan juga produk yang baik bagi pengguna/pembangun. Ia memudahkan produktiviti pangkalan data dari segi kebolegunaan dan kekonsistenan bagi Microsoft Windows. Oleh kerana Windows dan Access adalah dari produk Microsoft maka kedua-dua produk tersebut boleh saling bekerjasama. Microsoft Access boleh dilarikan pada Windows 95, 98 atau platform NT jadi kesemua kelebihan bagi Windows juga terdapat dalam Access. Walaubagaimanapun terdapat keterbatasan bagi Access dalam mengawal pangkalan data yang sangat besar. Jika diukur dari segi keefisienan, SQL Server adalah pilihan yang bijak.

4.4.3.3 MICROSOFT SQL SERVER 6.0

Microsoft SQL Server 6.0 menyediakan satu platform pengguna/pelayan bagi penghantaran aplikasi perniagaan kritikal bagi pelbagai persekitaran. Microsoft SQL Server 6.0 dibina sebagai berkuasa, kebolehskalaan dan kebolehdurusan oleh sistem pengendalian Microsoft Windows NT bagi menyediakan kebolehpercayaan dan keupayaan yang maju bagi pengurusan pangkalan data pelanggan/pelayan. Pada dasarnya ia dibangunkan bagi mengawal pangkalan data yang besar dan lebih cepat berbanding penyelesaian yang ada.

Microsoft SQL Server 6.0 menonjolkan perlaksanaan, kebolehpercayaan dan kebolehskalaan seperti yang disediakan oleh versi yang terlebih dahulu dikeluarkan bagi SQL Server dengan menjadikan proses bagi pembangunan aplikasi, pengurusan sistem dan penggantian data menjadi lebih mudah berbanding dahulu.

Microsoft SQL Server merupakan satu jenis pangkalan data perhubungan. Pangkalan data perhubungan adalah apabila data diurus ke dalam jadual antara satu pangkalan data. Jadual diurus oleh sekumpulan data tentang objek yang sama. Jadual tersebut akan dihubungkan semula antara satu sama lain oleh enjin carian pangkalan data apabila diperlukan.

SQL Server membenarkan pembangun sistem mengenalpasti sehingga 32,767 pangkalan data. Jumlah kesemua 2 juta jadual boleh dikenalpasti antara satu sama lain dalam 32,767 pangkalan data. Setiap jadual membenarkan

pengenalpastian sehingga 250 lajur. Ini bermakna SQL Server diurus bagi menyimpan sejumlah data yang sangat besar dan luas.[5]

4.4.3.4 VISUAL FOXPRO 6.0

Visual FoxPro dibangunkan daripada DOS- berasaskan bahasa Xbase kepada satu alatan pembangunan aplikasi visual GUI (Graphical User Interface) dalam masa yang agak singkat. Visual FoxPro memberikan implementasi yang memuaskan berorientasikan objek, kesesuaian-Xbase, bahasa pengaturcaraan pangkalan data dan sebagainya.

Di antara kelebihan Visual FoxPro adalah:

- a) Enjin carian data yang paling cepat
- b) Bahasa yang fleksibel untuk membina logik perniagaan kompleks menggunakan alatan pembangunan pangkalan data yang sebenar
- c) Sifatnya yang berorientasikan objek membolehkan ia membina rangka kerja serba guna yang mudah bagi pemprosesan permintaan
- d) Boleh menskalakan aplikasi melalui rangkaian

Di antara kelemahannya pula adalah:

- a) Ia adalah 'single-threaded', memerlukan pelbagai sesi Visual FoxPro bagi pemprosesan keserentakan
- b) Memerlukan skema penyelenggaraan yang berbeza [6]

4.4.3.5 ORACLE

Oracle merupakan peneraju vendor bagi perisian pangkalan data. Ia merupakan penyelesaian terbuka (open solution) dan akan menyokong semua jenis platform. Oracle berkebolehan mempunyai kesemua data dan dokumen yang disimpan dalam bilangan yang kecil bagi pelaksanaan pangkalan data yang tinggi.

Ini memberikan faedah kepada pengguna di mana Oracle akan memusatkan kesemua data, membuat pengurusan maklumat supaya mudah dicapai, boleh dipercayai dan tidak mahal. Tambahan pula, ciri keselamatan Oracle yang sedia ada membenarkan penguatkuasaan pelbagai keistimewaan, pengauditan yang pantas, menambahkan kawalan capaian, pemprosesan pengagihan yang selamat, pereplikaan (replication) dan kebolehan menggunakan mekanisme pengesahan tambahan luaran.[7]

4.4.3.6 PERBEZAAN ANTARA SQL SERVER DAN ORACLE

- a) SQL Server merupakan satu sistem yang lebih mudah digunakan berbanding dengan pangkalan data Oracle. Ia menyediakan banyak alatan grafik yang mesra pengguna bagi instalasi (installation), konfigurasi dan pentadbiran
- b) SQL Server hanya boleh dilarikan pada Windows tetapi Oracle mampu menyokong semua jenis platform. Oracle tidak bergantung pada mana-mana platform

- c) SQL Server telah disepadukan dengan produk-produk Microsoft yang lain

4.5 RUMUSAN

Keperluan sistem perlu dikaji dari pelbagai aspek bagi melancarkan proses pembangunan *CD4WAM*. Ini kerana setiap keperluan mempunyai perkaitan tersendiri antara satu sama lain. Justeru itu analisa terperinci perlu dilakukan bagi memilih keperluan perkakasan dan perisian yang bersesuaian dengan proses pembangunan *CD4WAM*. Peranan keperluan sistem adalah amat penting bagi memastikan objektif proses pembangunan *CD4WAM* tercapai.

REKABENTUK SISTEM

5.1 PENDAHULUAN

Rekabentuk adalah proses kreatif yang memerlukan pemahaman dan kebolehan semulajadi perekabentuk bagi menukarkan masalah kepada sesuatu bentuk penyelesaian. Ia adalah suatu proses di mana keperluan pengguna diubah kepada suatu persembahan dalam bentuk perisian. Persembahan ini akan memberi suatu gambaran menyeluruh tentang sistem.

Rekabentuk sistem memerlukan kreativiti pembangun sistem untuk menukarkan maklumat dan data yang diperolehi daripada fasa analisis kepada maklumat yang boleh diterima dan difahami oleh pengguna. Kepentingan rekabentuk sistem adalah dinyatakan dengan satu perkataan-kualiti. Rekabentuk sistem menyediakan pembangunan projek dengan perwakilan perisian yang boleh dinilai kualitinya. Rekabentuk sistem juga merupakan satu-satunya cara pembangun sistem menterjemah dengan tepat keperluan pengguna kepada satu sistem atau produk perisian yang telah siap dibangunkan. Tanpa rekabentuk sistem, pembangun sistem mungkin menghadapi risiko membina suatu sistem yang tidak stabil iaitu suatu sistem yang mungkin gagal berfungsi apabila sedikit perubahan dilakukan, suatu sistem yang mungkin sukar untuk diuji atau suatu sistem yang kualitinya tidak boleh dinilai sehingga akhir proses perisian, apabila masa adalah singkat dan apabila banyak wang yang telah dibelanjakan [8]

Rekabentuk sistem dilakukan secara manual dan lakaran tersebut membolehkan rekabentuk skrin yang sebenar pada fasa pengkodan mudah dilakukan. Walau bagaimanapun rekabentuk tersebut mungkin berubah dari semasa ke semasa mengikut daya kreativiti serta daya imaginasi pembangun sistem.

Rekabentuk sebenar terbahagi kepada dua proses iteratif iaitu **rekabentuk konsepsual** yang memberitahu pengguna apa yang sistem akan lakukan. Setelah pengguna atau pelanggan bersetuju dengan rekabentuk konsepsual ia akan diterjemahkan kepada dokumentasi yang lebih terperinci iaitu **rekabentuk logikal** yang membenarkan pembangun sistem memahami perisian dan perkakasan sebenar yang diperlukan untuk menyelesaikan masalah pengguna. Fasa rekabentuk akan dibahagikan kepada dua bahagian utama iaitu:

- a) Rekabentuk Struktur Sistem
- b) Rekabentuk Antaramuka Pengguna

5.2 REKABENTUK STRUKTUR SISTEM

Struktur sistem adalah alat yang digunakan untuk merekabentuk sistem ini. Ia merupakan interaksi dan hubungan antara modul-modul yang bergantung di antara satu sama lain. Rekabentuk struktur sistem boleh difahami dengan lebih jelas melalui rajah aliran proses.

5.2.1 ALIRAN PROSES

Aliran proses digunakan bagi menunjukkan setiap langkah yang terlibat dalam proses pembangunan *CD4WAM*. Setiap aktiviti ditunjukkan dalam aliran proses ini untuk memberi gambaran kepada pengguna tentang pelaksanaan *CD4WAM*. Rajah aliran proses bagi *CD4WAM* adalah seperti berikut:



TL Pekerja

Nama	T Lahir
Ali	1/6/1980
Abu	1/5/1990

Umur Pekerja

Nama	Umur
Ayub	30
Abu	0

Tarikh Lahir

Nama	Tarikh Lahir
Siti	0
Kasim	0

Umur

N Pekerja	Umur
Daud	11
Ramu	20



PANGKALAN DATA (SQL SERVER)

TL Pekerja

Nama	T Lahir
Ali	1/6/1980
Abu	1/5/1990

Umur Pekerja

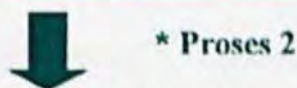
Nama	Umur
Ayub	30
Abu	0

Tarikh Lahir

Nama	Tarikh Lahir
Siti	0
Kasim	0

Umur

N Pekerja	Umur
Daud	11
Ramu	20



PANGKALAN DATA (SQL SERVER)

TL Pekerja

Nama	T Lahir
Ali	1/6/1980
Abu	1/5/1990
Kasim	UNKNOWN
Siti	UNKNOWN

Umur Pekerja

Nama	Umur
Ayub	30
Abu	UNKNOWN
Ramu	20
Daud	11

*** Proses 3**



GUDANG DATA

Proses 1 mewakili proses penggabungan antara dua pangkalan data yang berbeza format iaitu pangkalan data 1 adalah dalam format Microsoft Access manakala pangkalan data 2 adalah dalam format SQL Server. Kedua-dua pangkalan data ini akan digabungkan menjadi satu format piawai iaitu dalam format SQL Server. Data-data yang kotor dan nama bagi setiap entiti masih lagi dikekalkan.

Proses 2 mewakili proses pembersihan data. Jadual-jadual yang mengalami masalah akan dibersihkan menggunakan teknik nilai yang hilang. Tetapi teknik ini hanya akan membersihkan jadual-jadual yang mengalami nilai yang hilang sahaja. Kemudian jadual-jadual yang berkaitan akan disatukan bagi mengelakkan lewahan data. Nama bagi entiti juga diubah kepada nama yang lebih dominan. Akhirnya, pangkalan data yang telah dibersihkan akan disimpan dalam satu pangkalan data dalam format SQL Server.

Proses 3 mewakili proses memasukkan data-data yang telah dibersihkan ke dalam gudang data. Ini bagi memastikan gudang data hanya menyimpan data yang telah dibersihkan sahaja bagi memudahkan tugas pentadbir gudang data.

5.3 REKABENTUK ANTARAMUKA PENGGUNA

Antaramuka merupakan sebahagian daripada aplikasi yang pengguna boleh lihat dan berinteraksi dengannya. Ia adalah berkaitan dengan struktur, senibina dan juga pengkodan yang menjadikan sesuatu perisian itu berfungsi.

Antaramuka meliputi skrin, tettingkap(windows), kawalan, menu, metafora(metaphor), bantuan atas talian(on-line), dokumentasi dan juga latihan. Apa sahaja yang pengguna boleh lihat dan berkomunikasi dengannya merupakan sebahagian daripada antaramuka.

Rekabentuk antaramuka pengguna mencipta satu medium komunikasi yang efektif antara manusia dan komputer. Ia menuntut satu set prinsip-prinsip rekabentuk antaramuka pengguna di mana ia mengenalpasti objek dan tindakan antaramuka dan kemudiannya merekabentuk satu paparan skrin (screen layout) yang membentuk asas-asas bagi satu prototaip antaramuka pengguna. Oleh kerana antaramuka pengguna mempengaruhi persepsi pengguna pada sesuatu perisian ia perlu direka dengan baik. Walau sehebat manapun sesuatu perisian itu jika ia sukar digunakan menyebabkan pengguna melakukan kesilapan apabila menggunakannya, sudah tentu pengguna tidak berminat menggunakan perisian tersebut. Oleh itu rekabentuk antaramuka pengguna perlu dirancang dengan teliti bagi menghasilkan suatu sistem yang berfungsi dengan sempurna.

Bagi membangunkan rekabentuk antaramuka, pendekatan Penilaian Heuristik (heuristic evaluation) telah dijadikan sebagai panduan kepada pembangun sistem. Penilaian Heuristik dijalankan sebagai satu penyemakan yang sistematik ke atas rekabentuk antaramuka pengguna untuk menguji kebolehgunaannya. Matlamat utama Penilaian Heuristik adalah untuk mencari masalah kebolehgunaan dalam rekabentuk antaramuka.

5.3.1 10 HEURISTIK KEBOLEHGUNAAN

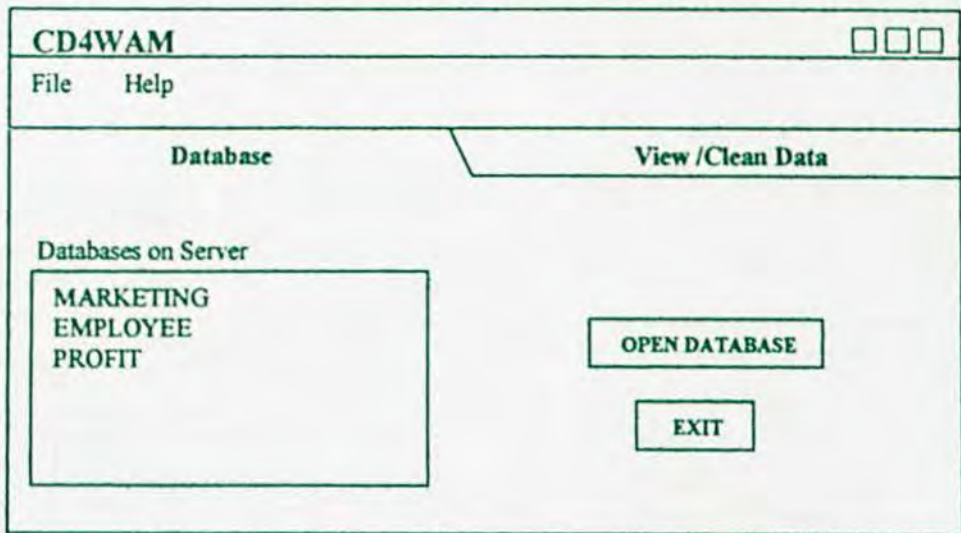
- a) Status sistem yang jelas
 - sistem tersebut hendaklah sentiasa memaklumkan pengguna mengenai apa yang sedang berlaku melalui maklumbalas yang sesuai dalam masa yang munasabah
- b) Bersesuaian dengan sistem dan dunia sebenar (real world)
 - sistem tersebut hendaklah menggunakan bahasa yang mudah difahami oleh pengguna. Perkataan, frasa dan konsep-konsep yang digunakan merupakan perkara yang biasa buat pengguna
- c) Kawalan dan kebebasan pengguna
 - pengguna selalunya tersilap memilih fungsi sistem dan memerlukan jalan keluar kecemasan (exit emergency) untuk meninggalkan keadaan yang tidak diingini itu tanpa perlu melalui dialog-dialog yang banyak dan panjang. 'Undo', 'Redo', Batal (Cancel), dan Kembali (Go Back) perlu disediakan bagi menangani masalah tersebut.
- d) Kekonsistenan dan kepiawaian
 - pengguna biasanya bergantung kepada kekonsistenan bagi memperolehi maklumat dengan cepat, mencipta model mental dengan tepat serta bagi membuat keputusan. Bagi mengelakkan pengguna daripada tertanya-tanya samada perkataan, situasi atau tindakan yang berlainan mempunyai maksud yang sama, adat platform (platforms conventions) perlu dipatuhi
- e) Pencegahan ralat

- rekabentuk yang teliti dapat mengelakkan masalah daripada timbul. Ini adalah lebih bagus daripada menghasilkan mesej ralat (error message)
- f) Pengiktirafan berbanding mengingat kembali (recognition rather than recall)
 - objek, tindakan dan pilihan dijadikan jelas dan nyata. Pengguna tidak perlu mengingati maklumat daripada satu bahagian ke satu bahagian yang lain. Arahan menggunakan sistem hendaklah jelas dan mudah diperolehi pada bila-bila masa.
- g) Penggunaan yang fleksibel dan efisien
 - sistem yang dihasilkan hendaklah berupaya melayan pengguna yang mahir dan kurang mahir. Pengguna dibenarkan bertindak dengan proses yang kerap berlaku dengan menyediakan 'short text' menggunakan sesuatu arahan. Ini akan dapat menjimatkan masa pengguna di samping mengelakkan kebosanan
- h) Rekabentuk yang minimal dan estetik
 - dialog mestilah mengandungi maklumat yang relevan atau sering digunakan. Ia juga hendaklah ringkas, padat dan tepat.
- i) Membantu pengguna mengenalpasti, menjalankan diagnostik dan memulih daripada ralat (recover from error)
 - mesej ralat hendaklah dinyatakan dalam bahasa yang mudah dan ringkas (tanpa kod) serta jelas menyatakan masalah dengan terperinci dan seterusnya mencadangkan penyelesaian masalah
- j) Bantuan (help) dan dokumentasi

- walaupun sistem adalah lebih bagus digunakan jika tiada dokumentasi namun begitu penyediaan bantuan dan dokumentasi mungkin diperlukan. Sebarang maklumat hendaklah mudah dicari, fokus kepada kerja pengguna, menyenaraikan langkah-langkah yang konkrit untuk dilaksanakan dan tidak terlalu besar.

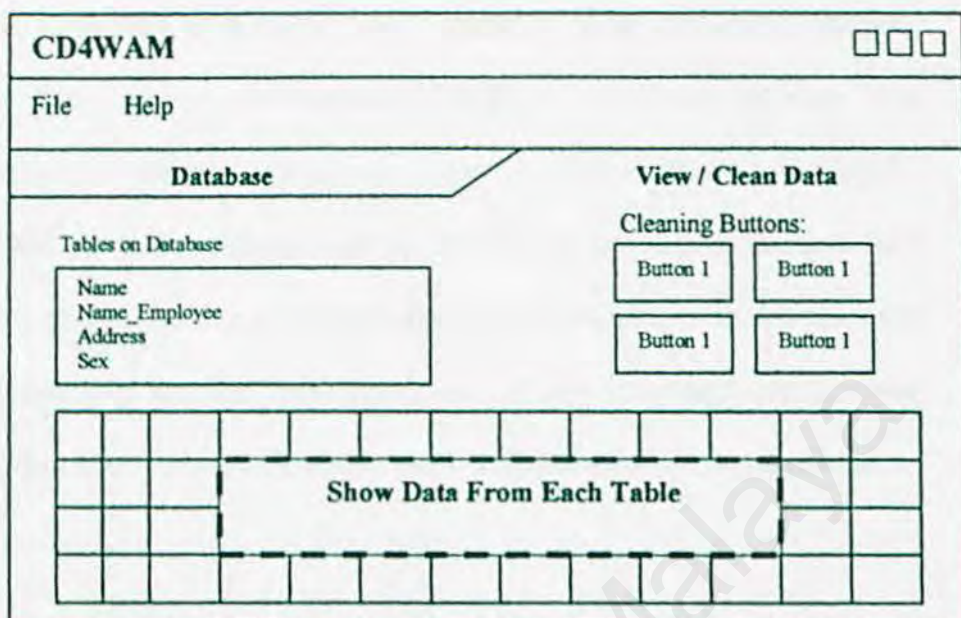
5.3.2 PERANCANGAN REKABENTUK ANTARAMUKA SISTEM

Antaramuka merupakan salah satu elemen penting bagi proses pembangunan *CD4WAM*. Antaramuka bagi *CD4WAM* tidak banyak melibatkan interaksi pengguna iaitu pentadbir gudang data kerana fokus utama *CD4WAM* adalah dari aspek pembersihan data. Oleh itu *CD4WAM* menyediakan satu antaramuka yang ringkas, mudah difahami dan dikendalikan oleh pentadbir gudang data. Rekabentuk antaramuka yang dicadangkan bagi *CD4WAM* adalah seperti rajah di berikut:



Rajah 5.3 Antaramuka 1

Antaramuka di atas merupakan antaramuka 1 bagi *CD4WAM* yang akan dibangunkan. 'Icon' *File* dan *Help* disediakan sebagai alatan tambahan. Pada 'icon' *File* terdapat arahan *Exit* di mana pengguna boleh keluar dari *CD4WAM*. 'Icon' *Help* menyediakan menu bantuan sekiranya pengguna *CD4WAM* menghadapi sebarang masalah. Terdapat dua 'tab' bagi antaramuka *CD4WAM* ini iaitu *Database* dan *View / Clean Data*. Pengguna perlu memasuki 'tab' *Database* terlebih dahulu. Kemudian pengguna perlu memilih salah satu pangkalan data yang terdapat pada pelayan untuk dibersihkan. Selepas itu pengguna perlu menekan butang *OPEN DATABASE* untuk memasuki pangkalan data yang dipilih. Butang *EXIT* adalah sebagai alternatif sekiranya pengguna ingin keluar dari *CD4WAM*.



Rajah 5.4 Antaramuka 2

Antaramuka 2 ini adalah antaramuka bagi 'tab' *View / Clean Data*. Ia akan memaparkan jadual-jadual yang terdapat dalam setiap pangkalan data. Ruang di bahagian bawah antaramuka akan memaparkan senarai atribut yang terdapat dalam sesebuah jadual dengan lebih terperinci. Bagi menjalankan proses pembersihan data, pengguna hanya perlu menekan butang yang disediakan disebelah kanan antaramuka. Setiap butang mewakili satu jadual yang ingin dibersihkan.

5.4 RUMUSAN

Rekabentuk antaramuka sistem merupakan salah satu elemen yang penting dalam proses pembangunan *CD4WAM*. Ini kerana ia akan turut mempengaruhi persepsi pengguna yang menggunakannya. Antaramuka *CD4WAM* yang akan dibangunkan adalah ringkas dan mudah difahami oleh pentadbir gudang data. Ia seharusnya memberi banyak pilihan kepada pentadbir gudang data bagi memilih pangkalan data-pangkalan data yang berkaitan untuk dibersihkan tanpa menimbulkan sebarang kekeliruan semasa menggunakannya.

IMPLEMENTASI DAN PERLAKSANAAN

6.1 PENDAHULUAN

Implementasi dan Pelaksanaan bagi *CD4WAM* dijalankan dengan merujuk kepada segala keperluan yang dicadangkan dalam Fasa Analisis dan Rekabentuk. Rujukan ini amat penting bagi memastikan pembangunan *CD4WAM* mematuhi segala keperluan yang perlu wujud dalam sesuatu proses pembangunan. Oleh yang demikian, perkara utama dalam fasa ini adalah bahagian pengaturcaraan yang mana merupakan senarai susunan set aturcara yang akan melarikan program.

Fasa ini melibatkan penterjemahan logik-logik setiap spesifikasi aturcara yang telah disediakan semasa Fasa Analisis dan Rekabentuk dalam bahasa-bahasa pengaturcaraan yang akan digunakan seperti Structured Query Language (SQL) serta VBscript. Proses-proses yang terlibat di dalam fasa ini adalah penghasilan unit-unit serta modul-modul aturcara yang dapat dikompilasikan dan boleh dilarikan dengan lancar termasuklah mewujudkan integrasi di antara antaramuka-antaramuka komponen sistem dengan pangkalan data.

6.2 PERSEKITARAN IMPLEMENTASI DAN PERLAKSANAAN

Di dalam Fasa Implementasi dan Pelaksanaan ini, penentuan persekitaran pembangunan *CD4WAM* adalah penting bagi memastikan proses pembangunan dapat dilaksanakan dengan lancar tanpa menimbulkan sebarang

masalah. Pemilihan perkakasan serta perisian yang betul dan sesuai dapat mempercepatkan proses pembangunan dan menjadi salah satu faktor kejayaan dalam pembangunan *CD4WAM*.

Perkakasan serta perisian yang digunakan dalam proses pembangunan *CD4WAM* adalah seperti berikut:

Perkakasan:

- kelajuan pemproses Intel Pentium 4 1.5 GHz atau seumpamanya
- papan kekeunci dan tetikus
- pemacu cakera keras dan pemacu cakera liut
- minimum 128 MB RAM

Perisian:

- Windows XP dan seumpamanya
- Microsoft SQL Server 7.0
- Microsoft Access 2000
- Visual Basic 6.0

6.3 FAKTOR-FAKTOR BERKAITAN

Sebelum fasa ini dilaksanakan beberapa faktor perlu diteliti. Ini bagi memastikan implementasi projek yang bakal dilaksanakan adalah sistematik dan bersesuaian. Antara faktor yang perlu diteliti ialah struktur proses *CD4WAM*

secara khusus dan keseluruhan. Setiap perkaitan atau integrasi antara satu komponen dengan yang lain akan dipastikan wujud di antaranya.

Komponen-komponen di dalam *CD4WAM* secara umumnya berkait antara satu sama lain. Sekiranya setiap perkaitan tidak dapat dihubungkan dengan jelas maka proses penghasilan *CD4WAM* akan terjejas. Ini kerana proses pembangunan *CD4WAM* tidak sama seperti proses pembangunan sistem-sistem yang lain. Sebagai contoh, sekiranya pangkalan data tidak dapat dikenalpasti di peringkat awal proses maka proses pembersihan pangkalan data tersebut tidak dapat dilakukan.

6.4 PENGATURCARAAN *CD4WAM*

Komponen-komponen yang terdapat di dalam *CD4WAM* dibentuk berasaskan persamaan logik, keperluan-keperluan data serta jujukan fungsi-fungsi yang berkenaan. Dua konsep yang digunakan di dalam implementasi dan pelaksanaan *CD4WAM* adalah konsep gandingan dan konsep ikatan. Konsep gandingan membentuk aturcara bermodul sementara konsep ikatan membentuk aturcara berstruktur.

6.4.1 PENGATURCARAAN BERMODUL

Pengaturcaraan bermodul merupakan kaedah pengaturcaraan yang membahagi suatu masalah yang kompleks kepada bahagian-bahagian yang lebih

kecil agar mudah diaturcarakan. Melalui kaedah ini kesilapan serta ralat dapat dikurang atau dielakkan sama sekali.

6.4.2 PENGATURCARAAN BERSTRUKTUR

Pengaturcaraan berstruktur pula merupakan kaedah pengaturcaraan secara bersistematik. Langkah-langkah yang dipraktikkan di dalam kaedah ini adalah:

- arahan-arahan cabang tanpa syarat perlu dielakkan atau diminimumkan penggunaannya dalam setiap modul aturcara
- arahan-arahan yang terkandung di dalam setiap rutin aturcara perlu berasaskan suatu jujukan logik agar ia mengandungi hanya satu punca kemasukkan ke dalam rutin dan satu punca keluar dari rutin
- setiap rutin mestilah mengandungi kod-kod lengkap dengan komen-komen yang mudah difahami

6.5 PENDEKATAN DALAM MENGIMPLEMENTASIKAN PENGATURCARAAN

Pengaturcaraan yang berkualiti seharusnya mempunyai ciri-ciri yang membantu ke arah pembinaan produk yang berkualiti iaitu mudah difahami, diimplementasi, diuji, diubahsuai dan bertepatan dengan segala keperluan yang telah dinyatakan. Dalam proses pengaturcaraan untuk membangunkan

CD4WAM, beberapa pendekatan pengaturcaraan diambil kira. Walaupun pendekatan yang digunakan tidak dipenuhi secara menyeluruh, namun konsep utama pendekatan ini telah digunakan sebagai panduan dalam pembangunan kod sumber. Konsep-konsep pengaturcaraan yang digunakan semasa pengaturcaraan ialah:

6.5.1 Pautan (Cohesion)

Pautan antara komponen adalah satu pengukuran terhadap sejauh mana perhubungan antara komponen-komponen tersebut. Satu komponen seharusnya melaksanakan satu fungsi logikal tertentu atau melaksanakan hanya 1 entiti logikal sahaja. Ia merupakan ciri-ciri unik kerana satu unit hanya mewakili satu bahagian dari penyelesaian masalah dan berpaut antara unit-unit yang lain. Oleh sebab itu, sekiranya ada perubahan yang perlu dibuat, pengaturcara hanya perlu mengubah unit-unit tertentu sahaja tanpa membuat perubahan pada keseluruhan aturcara sumber.

6.5.2 Kebolehfahaman (Understandibility)

Prinsip kebolehfahaman yang jelas pada rekabentuk dapat mengelakkan pengaturcara dari melakukan kesilapan dalam fasa implementasi. Di samping itu, dengan wujudnya kebolehfahaman yang tinggi, sebarang perubahan pada masa akan datang dapat dilakukan dengan mudah selain mampu mengelakkan kekeliruan dan kekompleksan pada aturcara.

6.5.3 Kebolehsuaian (Adaptibility)

Kebolehsuaian bagi rekabentuk adalah anggaran kasar bagaimana mudahnya perubahan dapat dilakukan kepada rekabentuk yang disediakan. Oleh sebab itu, komponen-komponen dalam aturcara sumber perlu dipaut supaya kebolehsuaian dapat dilakukan serentak tanpa melibatkan kesemua unit-unit.

6.6 PENGGUNAAN UTILITI DAN RUTIN SEPUNYA

Sesebuah sistem akan dapat dilaksanakan atau dibangunkan dalam masa yang singkat sekiranya proses pengaturcaraan dibangunkan secara cekap dengan menggunakan utiliti ataupun rutin sepunya yang sedia ada.

Sekiranya terdapat beberapa rutin yang sepunya pada kod-kod pengaturcaraan maka pengaturcaraan yang sama boleh digunakan semula untuk membentuk rutin sepunya dalam implementasi dan pelaksanaan *CD4WAM*. Cuma sekiranya terdapat beberapa fungsi tambahan yang perlu ditambah maka aturcara berkenaan perlu diubahsuai. Ini dapat menjimatkan masa dan tumpuan dapat diberikan kepada pembentukan aturcara yang lebih kompleks. Walaubagaimanapun kebanyakan aturcara komponen-komponen dalam *CD4WAM* adalah tidak sepunya dan tidak bersandar. Hanya terdapat beberapa fungsi yang sepunya. Maka aturcara bagi setiap komponen secara amnya adalah berbeza di antara satu sama lain.

6.7 PERLAKSANAAN PROSES PENGATURCARAAN

Di dalam pelaksanaan proses pengaturcaraan *CD4WAM*, beberapa langkah perlu dituruti seperti penyediaan pengkodan aturcara, pengkodan setiap komponen aturcara, pengujian setiap komponen aturcara yang telah dikodkan, pengujian keseluruhan *CD4WAM* serta mendokumentasikan aturcara-aturcara yang telah dibentuk atau dibangunkan. Langkah-langkah tersebut adalah sebagai garis panduan agar pelaksanaan proses pengaturcaraan dapat dilaksanakan secara sistematik dan lancar.

Berikut adalah langkah-langkah berkenaan proses pengaturcaraan secara am:

- spesifikasi pengkodan
- pengkodan aturcara
- kompilasi dan himpunan aturcara

6.8 HASIL FASA IMPLEMENTASI DAN PERLAKSANAAN

CD4WAM [] [] []

File Settings Help

Login to server / **View / Clean Data**

Databases on Server

MARKETING EMPLOYEE PROFIT

UserName: john

PassWord: *****

ServerName: network

OPEN DATABASE CLOSE

CLOSED READ

Rajah 5.5 Antaramuka 1 yang telah dipilih

CD4WAM [] [] []

File Settings Help

Login to server / **View / Clean Data**

Tables on Database

AKAUN AKAUN_PELGN CAWANGAN CAW_SYARIKAT
--

Columns in Table

NoAkaun Nama Alamat NoTel Jantina

Cleaning Process:

AKAUN	Atribut 2
Atribut 3	Atribut 4

NoAkaun	Nama	Alamat	NoTel	Jantina
235-255-47788	SITI HAJAR	PEKAN,PAHANG	09-58774145	P
561-210-45770	HAIKAL	MIJAR JOHOR	07-51124410	I
455-201-47201	SUHANA	AMPANG	03-55474122	

OPEN READ Local SQL-Server Found: 4 Objects EMPLOYEE

Rajah 5.6 Antaramuka 2 yang telah dipilih

6.9 RUMUSAN

Bagi mengaplikasikan teknik nilai yang hilang dalam proses pembersihan data, pembangun telah memilih untuk menggunakan dua kaedah iaitu menggantikan nilai yang hilang dengan dua pembolehubah global iaitu 'UNKNOWN' dan '0' serta menggunakan atribut bermakna. Pemilihan kaedah yang tepat adalah bergantung kepada pangkalan data yang terlibat. Antaramuka yang dipilih adalah sama seperti yang dinyatakan dalam Fasa Analisis dan Rekabentuk. Cuma terdapat sedikit penambahan dari segi paparan data dan kawalan. Ini kerana proses *CD4WAM* adalah sukar untuk dibayangkan secara nyata dan diharapkan agar antaramuka yang dipilih mampu membantu pentadbir gudang data memahami aliran proses yang terlibat dalam *CD4WAM*. Tambahan pula pangkalan data yang terlibat hanya dapat dicapai melalui pelayan memberikan sedikit cabaran kepada pembangun untuk menghubungkan setiap komponen yang terlibat.

PENGUJIAN DAN PENYELENGGARAAN

7.1 PENDAHULUAN

Pengujian dan penyelenggaraan merupakan fasa terakhir di dalam proses pembangunan *CD4WAM*. Pengujian merupakan aspek penting bagi menentukan tahap kualiti sesuatu sistem yang bakal menepati spesifikasi yang telah ditetapkan dan memenuhi keperluan dan kehendak pengguna.

Selaras dengan objektif pencapaian spesifikasi yang telah ditetapkan, beberapa objektif lain di dalam fasa pengujian dan penyelenggaraan juga dapat dicapai iaitu:

- mengenalpasti dan memperbaiki ralat yang wujud
- melaksanakan pengujian terhadap komponen-komponen sistem yang berkaitan bagi memastikan komponen-komponen berkenaan berfungsi dengan lancar serta menepati keperluan dan kehendak pengguna

CD4WAM

7.2 PERINGKAT PENGUJIAN

Terdapat tiga jenis pengujian yang dilakukan ke atas *CD4WAM* iaitu pengujian unit, pengujian integrasi dan juga pengujian sistem

7.2.1 PENGUJIAN UNIT

Pengujian unit juga dikenali sebagai pengujian modul di mana ia dijalankan mengikut modul secara berasingan. Ini adalah bagi memastikan setiap modul di dalam *CD4WAM* dapat dijalankan dengan sempurna tanpa sebarang ralat. Semasa pengujian unit, setiap ralat yang wujud pada setiap modul akan dikesan dan dihapuskan.

7.2.2 PENGUJIAN INTEGRASI

Setelah modul-modul dalam *CD4WAM* dipastikan dapat berfungsi dengan sempurna serta menepati fungsi masing-masing, modul-modul ini akan diintegrasikan agar membentuk sistem. Integrasi ini dirancang dan dikoordinasikan dengan betul supaya apabila berlaku kesilapan, pengaturcara dapat mengesan ralat tersebut. Ujian secara keseluruhan dilakukan bagi melihat setiap hubungan dapat dilaksanakan dengan baik. Di sini tumpuan diberikan ke atas rekabentuk antaramuka setiap bahagian yang digabungkan.

7.2.3 PENGUJIAN SISTEM

Pengujian sistem adalah pengujian yang terakhir. Ia bermula apabila aturcara-aturcara yang telah berjaya dilarikan dalam ujian integrasi. Pengujian sistem melibatkan pengujian dengan masalah sebenar dan ianya melibatkan

pengguna. Pengujian sistem memeriksa samada elemen sistem dapat berfungsi dengan betul dan menepati objektifnya.

Di antara objektif-objektif pengujian sistem adalah untuk:

- mengukur prestasi, kelemahan dan keupayaan sistem secara keseluruhannya samada ianya dapat mencapai tahap yang boleh diterima
- mengesahkan ketepatan dan kejitian semua komponen sistem yang dibangunkan berdasarkan spesifikasi-spesifikasi subsistem yang telah direkabentuk. Setiap subsistem dipastikan agar boleh dilarikan dengan lancar dan akan berfungsi sebagaimana yang dikehendaki dalam keadaan yang serupa dengan persekitaran operasi yang sebenar
- mengukur sejauh mana sistem yang dibangunkan itu dapat memenuhi objektif-objektif yang telah ditentukan

Selain itu kesesuaian antaramuka *CD4WAM* dengan persekitaran sistem pengendalian yang digunakan turut dinilai. Ini adalah untuk menguji samada ianya dapat dilarikan sebagai sebuah program atau tidak. Ujian pada peringkat ini dipanggil sebagai ujian kotak hitam. Kemudian pengujian program dilakukan untuk mencari kelemahan dan mengukur keupayaan serta menguji kesepaduan program dan mengesahkan ianya memenuhi keperluan yang telah ditetapkan.

Bagi *CD4WAM*, perkara-perkara yang dinilai adalah:

- pemetaan skrin - pengurusan menu dan antaramuka pengguna
- pengujian dokumentasi - menguji semua contoh dalam manual pengguna

- pengujian bantuan program

Terdapat lima jenis asas ujian tambahan dilakukan kepada ujian fungsian iaitu:

- kebolegunaan - dari segi antaramuka yang biasa digunakan oleh pengguna
- kebolehpercayaan - ujian kelakuan sistem
- prestasi - masa tindakbalas
- kebolehselenggaraan - kesediaan untuk membetulkan sebarang kerosakkan

Pengujian sistem melibatkan:

- ujian fungsi - ujian ini memeriksa sistem yang telah diintegrasikan menjalankan fungsinya seperti yang ditetapkan oleh keperluan sistem
- ujian pencapaian - ujian untuk membandingkan komponen yang telah diintegrasikan dengan keperluan sistem bukan fungsian. Antaramuka sistem diuji semasa fasa ini
- ujian penerimaan - ujian bagi memastikan sistem beroperasi seperti yang dikehendaki oleh pembangun dan pengguna. Dalam ujian ini, pengguna dikehendaki menggunakan program tanpa sebarang bantuan daripada pembangun. Ia bertujuan untuk menguji kefahaman pengguna ke atas setiap rekabentuk antaramuka pengguna dan fungsinya.

Pengujian ke atas *CD4WAM* dilaksanakan apabila proses pengujian integrasi sistem selesai dan berjalan lancar tanpa sebarang ralat. Objektif pengujian sistem dilaksanakan adalah untuk memastikan:

- *CD4WAM* mencapai ketepatan serta kejituan kesemua komponen yang dibangunkan di dalam sistem mengikut spesifikasi yang telah ditetapkan dan yang paling penting ianya memenuhi keperluan dan kehendak pengguna sistem. Setiap subsistem akan dipastikan beroperasi dengan lancar sehingga keseluruhan sistem dapat dilarikan tanpa sebarang masalah. *CD4WAM* yang telah direkabentuk dipastikan dapat dilarikan dalam persekitaran pengoperasian yang sebenar
- mengukur dan menilai pengoperasian *CD4WAM* secara keseluruhan samada dapat dilarikan pada tahap yang boleh diterima
- membandingkan *CD4WAM* yang telah dibangunkan dengan objektif-objektif atau spesifikasi yang perlu dicapai. Ini dapat mengesan sebarang kekurangan pada pengoperasian sistem pada keseluruhannya

Data-data 'input' telah ditentukan untuk menguji kelancaran sistem dan hasilnya diperhatikan. Ini juga dapat menguji kebolehlaksanaan integrasi *CD4WAM* dengan setiap pangkalan data serta integrasi dengan antaramuka-antaramuka komponen sistem.

7.3 PENYELENGGARAAN *CD4WAM*

Pada bahagian ini tidak banyak yang dapat dilakukan ke atas *CD4WAM*. Ini kerana *CD4WAM* bukannya satu sistem yang memerlukan data yang sentiasa dikemaskini dan berubah-ubah. Penyelenggaraan yang dilakukan adalah lebih difokuskan kepada kod-kod aturcara pembersihan pangkalan data. Sekiranya kod-kod aturcara tidak berjalan seperti yang dikehendaki oleh pembangun maka ianya perlu diubahsuai.

7.4 KEPERLUAN PENYELENGGARAAN

Penyelenggaraan sistem adalah penting bagi memastikan *CD4WAM* yang telah dibangunkan dapat terus beroperasi atau berfungsi dari masa ke semasa tanpa sebarang ralat atau kesilapan. Ini kerana penyelenggaraan sistem akan memastikan pengoperasian *CD4WAM* sentiasa dalam keadaan lancar setiap masa. Aktiviti penyelenggaraan mungkin dilakukan dalam skala mingguan atau bulanan mengikut saranan pengguna sistem. Sekiranya sebarang masalah timbul, ianya akan diatasi dengan segera dan sistem akan diselenggara sehingga sistem beroperasi seperti sedia kala. Ini bagi memastikan persekitaran pengoperasian *CD4WAM* sentiasa menepati spesifikasi serta keperluan dan kehendak pengguna sistem.

7.5 RUMUSAN

Fasa pengujian dan penyelenggaraan adalah satu fasa yang amat penting agar pembangun dapat menganalisa sejauh masa kejayaan penghasilan *CD4WAM* serta memastikan samada objektif yang telah ditetapkan tercapai. Hasil yang diperolehi dari fasa pengujian akan dikenalpasti dan diperbaiki sepenuhnya dalam fasa penyelenggaraan bagi memastikan operasi keseluruhan *CD4WAM* berjalan seperti yang dikehendaki oleh semua pihak yang terlibat.

PENILAIAN SISTEM

8.1 PENDAHULUAN

Fasa ini melibatkan pengguna *CD4WAM* untuk menilai sistem secara keseluruhan samada menepati kesemua keperluan dan kehendak pengguna. Di samping itu fasa ini juga akan memfokus kepada masalah-masalah yang timbul semasa pembangunan serta cara penyelesaiannya, kelebihan sistem, keterbatasan sistem serta cadangan untuk meningkatkan prestasi *CD4WAM* pada masa akan datang. Objektif utama fasa penilaian sistem adalah:

- membuat penilaian terhadap keseluruhan sistem oleh pengguna dan pembangun *CD4WAM*. Ini bagi memastikan kesemua fungsi atau pelaksanaan sistem yang tidak menepati keperluan dan kehendak pengguna dan pembangun dapat diperbaiki dan dipertingkatkan

8.2 MASALAH DAN PENYELESAIAN

Semasa proses membangunkan *CD4WAM* terdapat beberapa kesukaran serta masalah yang dihadapi oleh pembangun terutamanya berkaitan kod aturcara bagi menjanakan output seperti yang diinginkan. Walaubagaimanapun kesukaran serta masalah yang timbul dapat diatasi dengan mengambil langkah penyelesaian seperti yang dinyatakan dibawah:

i) **kurang penguasaan dalam pengaturcaraan**

- tahap penguasaan pembangun dalam pengaturcaraan adalah agak lemah kerana kurangnya pendedahan sebelum ini. Maka pembangun menghadapi kesukaran di dalam menyediakan aturcara yang sesuai dan tepat untuk menghasilkan sesuatu fungsi sistem seperti yang dikehendaki. Antara bahasa pengaturcaraan yang digunakan adalah Structured Query Language (SQL) dan VBScript.

Langkah penyelesaian:

- Pembangun telah memperuntukkan masa yang lebih untuk mempelajari bahasa-bahasa pengaturcaraan tersebut sebelum memulakan proses membangunkan *CD4WAM*. Seiring dengan proses membangunkan *CD4WAM*, pembangun telah terdedah dengan konsep serta cara pengaturcaraan yang lebih baik merujuk kepada bahasa pengaturcaraan yang digunakan. Walaupun cara ini mengambil masa dan mungkin mengganggu penjadualan projek yang telah ditetapkan tetapi cara ini adalah terbaik dan berkesan kepada pembangun.

ii) **kurang pendedahan terhadap aspek-aspek lain**

- selain itu pembangun juga menghadapi kesukaran di dalam aspek-aspek seperti sistem pangkalan data, konsep pelayan, integrasi antaramuka pengguna yang bersesuaian dan sebagainya. Ini kerana pembangun

kurang terdedah dan kurang berpengalaman di dalam aspek-aspek tersebut.

Langkah penyelesaian:

- Pembangun telah mengambil inisiatif dengan mempelajari aspek-aspek tersebut samada melalui buku rujukan, melayari laman web yang berkaitan serta berbincang dengan rakan-rakan yang lebih mahir berkenaan aspek-aspek tersebut. Ini menambahkan pengetahuan pembangun serta memberikan gambaran yang lebih jelas terhadap aspek-aspek yang digunakan di dalam proses membangunkan *CD4WAM*. Oleh itu inisiatif yang berterusan perlu dilakukan untuk meluaskan pengetahuan pembangun dari semasa ke semasa.

iii) kurang kemahiran teknikal

- satu lagi kesukaran yang dihadapi oleh pembangun adalah kurangnya kemahiran secara teknikal. Sebelum ini di dalam kursus-kursus yug diambil kebanyakannya memberi pendedahan dan penekanan terhadap konsep-konsep secara teori. Maka pengetahuan pembangun di dalam bidang teknikal adalah kurang. Masalah timbul berkaitan dengan pengetahuan teknikal contohnya apabila berlaku kerosakkan pada perkakasan dan perisian komputer, serangan virus pada komputer dan sebagainya.

Langkah penyelesaian:

- Sekiranya berlaku sebarang masalah atau kesukaran berkaitan masalah teknikal yang tidak dapat diatasi, nasihat serta bantuan daripada mereka yang berkemahiran adalah diperlukan terutamanya rakan-rakan. Pengetahuan yang diperolehi daripada masalah yang dihadapi dapat dijadikan panduan sekiranya masalah yang sama timbul kelak.

iv) peruntukkan masa pembangunan yang singkat

- oleh kerana kurangnya pendedahan dan pengalaman di dalam membangunkan sebuah projek sebelum ini, menyebabkan lebih masa diperlukan untuk mempelajari konsep-konsep yang berkaitan dan seterusnya membangunkan sistem. Ini menyebabkan peruntukkan waktu bagi proses pembangunan sistem menjadi singkat.

v) kurang sumber rujukan

- disebabkan *CD4WAM* adalah merupakan projek yang baru dibangunkan maka ini menimbulkan masalah kepada pembangun di mana sukar mencari rujukan bagi proses penghasilannya. Pembangun sukar merealisasikan *CD4WAM* dalam dunia sebenar. Proses penghasilannya *CD4WAM* agak rumit untuk difahami kerana ianya banyak melibatkan proses aturcara pada bahagian 'back-end'.

Langkah penyelesaian:

- Pembangun banyak mengadakan perbincangan bersama penyelia dan rakan-rakan yang berpengetahuan tinggi berkaitan sistem pangkalan data. Ini ditambah dengan pencarian maklumat-maklumat berkaitan yang terdapat di internet.

vi) tidak mendapat kerjasama dari organisasi yang terlibat

- Pembangun telah pergi ke beberapa organisasi bagi mendapatkan maklumat dan pangkalan data yang bersesuaian. Malangnya organisasi tersebut tidak memberikan kerjasama kerana perkara tersebut banyak berkaitan dengan data-data rahsia organisasi.

Langkah penyelesaian:

- Pembangun telah mengadakan perbincangan bersama rakan-rakan bagi merekabentuk pangkalan data yang bersamaan dengan masalah yang ingin diselesai. Mungkin pangkalan data yang telah dihasilkan tidak sama seperti pangkalan data dalam dunia sebenar tetapi apa yang penting sekali ialah teknik yang ingin diguna berjaya diaplikasikan.

8.3 KELEBIHAN *CD4WAM*

Di antara kelebihan-kelebihan *CD4WAM* adalah seperti berikut:

- i) pengguna *CD4WAM* boleh menggunakan 'default' pelayan yang terdapat pada 'icon' *Settings* sekiranya pelayan yang dipasang pada komputer atau organisasi tidak menggunakan 'user name' dan 'password'.
- ii) *CD4WAM* boleh memaparkan kesemua pangkalan data yang terdapat dalam pelayan
- iii) *CD4WAM* mampu memaparkan kesemua jadual dan atribut yang wujud dalam setiap pangkalan data samada secara logikal mahupun fizikal
- iv) pentadbir gudang data hanya perlu menekan butang yang disediakan bagi menjalankan proses pembersihan data bagi sesuatu jadual
- v) pengguna baru boleh mempelajari menggunakan *CD4WAM* melalui 'icon' *Help* yang disediakan

8.4 KETERBATASAN *CD4WAM*

Di sebalik kelebihan-kelebihan yang telah dinyatakan sebelum ini, *CD4WAM* turut mempunyai keterbatasan tertentu seperti:

- i) *CD4WAM* hanya terhad bagi pembersihan pangkalan data-pangkalan data yang telah ditetapkan sahaja. Sekiranya *CD4WAM* ingin diaplikasikan pada pangkalan data yang lain, kod aturcara bagi proses pembersihan data perlu diubahsuai
- ii) pangkalan data-pangkalan data yang ingin dibersihkan mestilah berada dalam pelayan Microsoft SQL Server

- iii) pangkalan data dalam format lain seperti Access, Oracle, Paradox dan sebagainya tidak boleh digunakan. Sekiranya pangkalan data dalam format lain hendak digunakan, pangkalan data tersebut haruslah dipindahkan ke dalam format Microsoft SQL Server
- iv) kaedah bagi pembersihan data menggunakan teknik nilai yang hilang adalah berbeza-beza mengikut data yang terdapat dalam pangkalan data

8.5 PENINGKATAN *CD4WAM* DI MASA HADAPAN

Beberapa fungsi atau komponen *CD4WAM* boleh dikemaskini atau diperbaharui untuk aplikasi yang terbaik pada masa hadapan. Kemungkinan pembangun sistem pada masa hadapan boleh memikirkan fungsi-fungsi atau komponen-komponen yang lebih baik dan efisien. Terutamanya dengan memperbaiki atau memperkemaskini keterbatasan *CD4WAM* yang telah dinyatakan sebelum ini. Berikut adalah beberapa cadangan untuk meningkatkan prestasi *CD4WAM* secara keseluruhan pada masa hadapan:

- i) *CD4WAM* boleh digunakan pada mana-mana pangkalan data yang ingin dibersihkan dan tidak hanya tertumpu pada pangkalan data yang telah ditetapkan
- ii) *CD4WAM* mampu melakukan proses pembersihan pangkalan data dari pelbagai format dan tidak menumpu kepada satu format sahaja
- iii) *CD4WAM* mampu mengaplikasikan kesemua kaedah bagi proses pembersihan pangkalan data menggunakan teknik nilai yang hilang

- iv) kaedah pembersihan data seharusnya selaras bagi semua atribut yang terlibat

8.6 CADANGAN

Di sini pembangun ingin mencadangkan beberapa pendapat yang mungkin dapat diatasi dan diambil perhatian oleh semua pihak:

- i) Pihak Fakulti Sains Komputer dan Teknologi Maklumat boleh menyediakan lebih banyak komputer khas untuk para pelajar melaksanakan Projek Ilmiah Tahap Akhir II. Ini sebagai langkah kontigensi bagi memastikan setiap pelajar dapat menggunakan segala kemudahan komputer yang disediakan sepenuhnya dan menyiapkan projek dalam masa yang ditetapkan.
- ii) Pihak Fakulti Sains Komputer dan Teknologi Maklumat boleh menyediakan kemudahan perisian yang diperlukan oleh para pelajar di dalam pelaksanaan projek. Ini memudahkan para pelajar membangunkan sistem dengan kemudahan perisian yang tersedia ada tanpa perlu mengeluarkan kos yang lebih untuk mendapatkan keperluan perisian berkenaan
- iii) Pihak Fakulti Sains Komputer dan Teknologi Maklumat boleh menyediakan lebih banyak rujukan seperti Laporan Projek Ilmiah Tahap Akhir pelajar-pelajar terdahulu serta buku-buku rujukan berkaitan pembangunan sistem serta pembelajaran penggunaan perisian yang bersesuaian. Diharap juga agar

pihak Fakulti bukan sahaja dapat menyediakan 'hard-copy' tetapi juga 'soft-copy' bagi Projek Ilmiah 2 pelajar-pelajar terdahulu. Ini bukan sahaja dapat dijadikan rujukan oleh para pelajar malah mampu menjana idea baru bagi projek-projek yang bakal dihasilkan. Jadual waktu penggunaan Bilik Dokumen juga diharap dapat dipanjangkan tempohnya dan sentiasa dibuka mengikut jadual yang ditetapkan.

- iv) Pihak Perpustakaan Utama Universiti Malaya juga boleh memainkan peranan dengan memperbanyakkan buku-buku rujukan terkini berkaitan dunia komputer. Ini dapat membantu para pelajar untuk mendalami pengetahuan berkaitan perisian, perkakasan atau aspek-aspek lain berkenaan dunia komputer.

8.7 PENGAJARAN DAN PENGALAMAN

Secara keseluruhannya, pelbagai pengajaran serta pengalaman yang diperolehi oleh pembangun dan boleh dijadikan panduan dan iktibar di masa hadapan. Di antara pengajaran dan pengalaman yang diperolehi semasa membangunkan *CD4WAM* adalah:

- i) mempertingkatkan lagi tahap penguasaan pembangun di dalam bahasa-bahasa pengaturcaraan seperti Sturctured Query Language (SQL), VBScript serta sistem pangkalan data dalam format yang lebih maju seperti Oracle, Paradox dan sebagainya

- ii) mempertingkatkan pengetahuan di dalam aspek-aspek pembangunan sesebuah sistem seperti aspek pelayan/pelanggan, integrasi sistem, sistem pangkalan data dan sebagainya
- iii) terdedah kepada kemahiran teknikal berkaitan perkakasan dan perisian
- iv) memperolehi pengalaman di dalam pembangunan sistem dalam situasi dan persekitaran sebenar. Di samping mempelajari fasa-fasa yang terlibat di dalam pelaksanaan projek serta pembangunan sistem. Juga menjadualkan proses pembangunan projek dengan lebih konsisten
- v) menimba pengalaman tentang cara-cara mengatasi masalah dengan baik serta melaksanakan langkah berjaga-jaga bagi menghindari sesuatu masalah atau kesukaran daripada berlaku
- vi) mendisiplinkan diri dengan sikap-sikap seperti sabar, bertanggungjawab, berdedikasi, yakin diri, menepati waktu serta tidak berputus asa di dalam melaksanakan sesuatu projek

8.8 KESIMPULAN

Secara keseluruhannya *CD4WAM* berperanan sebagai pembantu kepada pentadbir gudang data bagi menjalankan proses pembersihan data di dalam pangkalan data menggunakan teknik nilai yang hilang dan kemudiannya dimasukkan ke dalam gudang data. Pengguna *CD4WAM* hanya tertumpu kepada pentadbir gudang data sahaja. Sehingga proses akhir fasa pembangunan *CD4WAM*, pembangun mendapati objektif utama *CD4WAM* telah berjaya dicapai tetapi masih banyak lagi keterbatasan yang ada pada *CD4WAM*. Ini kerana projek sebegini merupakan projek yang pertama dibangunkan maka ini merupakan satu cabaran dan masalah yang besar kepada pembangun dalam membangunkan *CD4WAM* iaitu untuk merealisasikannya dengan dunia sebenar. Tambahan pula kurangnya rujukan di internet dan sistem sedia ada menyebabkan proses pembangunan *CD4WAM* bergerak perlahan dan sampai kepada tahap pembangun rasa berputus asa. Tidak ramai rakan-rakan dan kenalan pembangun yang mahir berkaitan sistem pengurusan pangkalan data dan ini menyukarkan pembangun untuk meminta pertolongan.

Sepanjang tempoh pembangunan dan pelaksanaan *CD4WAM*, banyak perkara-perkara baru yang dipelajari oleh pembangun. Ia bukan sahaja dari aspek pengaturcaraan perisian malah meliputi perkakasan dan segala aspek yang berkaitan dengan komputer.

Kesimpulannya, adalah diharapkan agar *CD4WAM* mampu menjana idea baru dalam sistem pengurusan pangkalan data di masa akan datang kerana

bidang ini masih baru untuk diterokai. Akhir kata diharapkan juga agar *CD4WAM* ini akan dapat diambil perhatian para pensyarah dan pelajar untuk diaplikasikan dan dimajukan pada masa hadapan mengikut keperluan semasa.

University of Malaya

BIBLIOGRAFI

- [1] George M. Maracas (1999). *Decision Support System In The Twenty-First Century*. Prentice-Hall.
- [2] Jiawei Han, Micheline Kamber (2002). *Data Mining: Concepts And Techniques*. Morgan Kaufmann Publishers.
- [3] Shari Lawrence Pgleeger (2002). *Software Engineering Theory And Practice*. Second Edition. Prentice Hall.
- [4] Howe D (1995). *The Free On-line Dictionary of Computing*.
<http://wombat.doc.ic.uk>
- [5] My SQL Homepage
<http://www.mysql.com/documentation>
- [6] Basian. M, Booth, J (1998). *Using Visual FoxPro 6.0, Special Edition*. 2nd Edition. Sams Publishing.
- [7] Michael Richards (1996). *Oracle? Unleashed*. 1st Edition. Sams Publishing.
- [8] Pressman, Rogers (2001). *Software Engineering : A Practitioner's Approach*. 5th Edition. Mcgraw Hill.
- [9] Dr. Abdullah Embong (2000). *Sistem Pangkalan Data: Konsep Asas, Rekabentuk Dan Perlaksanaan*. Tradisi Ilmu Sdn. Bhd.
- [10] H.M. Deitel, P.J. Deitel, T.R. Neito (2001). *Internet & World Wide Web: How To Program*. 2th Edition. Prentice Hall.
- [11] Shelly, Cashman, Repede, Mick (2001). *Microsoft Visual Basic 6.0: Complete Concepts And Techniques*. Course Technology.
- [12] Tim Anderson (2002). *Visual Basic In Easy Steps*. Computer Step.
- [13] <http://www.planet-source-code.com>
- [14] <http://www.vbscript.com>
- [15] <http://www.3Dtextmaker.com>
- [16] <http://www.w3schools.com>

- [17] Stephen Wynhoop (1999). *Special Edition using Microsoft SQL Server 7.0*. Qve Corporation.
- [18] Anthony T.Mann (1999). *Microsoft SQL Server 7.0 For Dummies*. IDG Books Worldwide, INC.
- [19] Thomas M.Connolly, Carolyn Begg, Anne D.Strachan (1999). *Database Systems*. Addison Wesley.

University of Malaya