



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Gambaran Umum Instansi

Universitas Multimedia Nusantara (UMN) adalah perguruan tinggi swasta yang didirikan oleh Kelompok Kompas Gramedia pada tanggal 20 November 2006 yang terletak di Scientia Garden, Gading Serpong, Tangerang. UMN memiliki empat fakultas dan program studi yaitu:

- a. Fakultas Desain Komunikasi Visual (DKV) atau *Visual Communication Design*

Fakultas yang memiliki fokus di bidang seni dengan menggunakan teknologi. DKV memiliki tiga program peminatan:

- i. Desain Grafis (DG) atau *Graphic Design*.
- ii. Animasi atau *Animation*.
- iii. Sinematografi atau *Digital Cinematography*.

- b. Fakultas Ilmu Komunikasi (Komunikasi Multimedia atau *Multimedia Communication*)

Fakultas yang memiliki fokus pada komunikasi menggunakan media digital. Ilmu komunikasi memiliki dua program peminatan:

- i. *Journalism*
- ii. *Public Relation (PR)*

c. Fakultas Teknologi Informasi dan Komunikasi atau *Information Technology and Communication (ICT)*

Fakultas yang fokus di bidang teknologi, informasi, dan komunikasi.

ICT memiliki tiga program studi:

- i. Program Studi Teknik Informatika (*Mobile Software Engineering*).
- ii. Program Studi Sistem Komputer (*Computer Engineering*).
- iii. Program Studi Sistem Informasi (*Mobile Information System*).

d. Fakultas Ekonomi

Fakultas yang fokus di bidang keuangan, bisnis dan studi mengenai ekonomi lainnya. Ekonomi memiliki dua program studi:

- i. Program Studi Manajemen (*ICT Business Management*)  
Manajemen memiliki dua peminatan, yakni:
  - (a). Manajemen Pemasaran (*Marketing Management*)
  - (b). Manajemen Keuangan (*Finance Management*)

- ii. Program Studi Akuntansi (*ICT Based Accounting*)

Peminatan pada akuntansi:

- (a). Auditing.
- (b). Perpajakan atau *Taxation*.
- (c). Sistem Informasi Akuntansi atau *Accounting Information System*.

Seluruh fakultas di UMN menggunakan multimedia dan teknologi dalam perkuliahan sehingga mahasiswa dituntut untuk memiliki kemampuan bahasa Inggris dan matematika yang baik.

### 3.2 Gambaran Umum Riset

Pada riset ini terbagi menjadi tiga aktivitas utama yakni pengumpulan dan pergudangan data (*data collection & data mart*) yang dikelola oleh penulis, dimulai dari hasil pengumpulan data hingga tampilan tabel fakta *data mart* yang dialokasikan ke penambangan data (*data mining*) yang dikelola oleh Fededi (11110310029), dari penerimaan data dari *data mart* hingga hasil analisis untuk mengetahui kesenjangan di dalam kemampuan bahasa Inggris per daerah. Terakhir adalah visualisasi data (*data visualization*) yang dikelola oleh Wisnu Satyagraha (11110310008), dimulai dari hasil analisis hingga hasil akhir data yang ditampilkan melalui Google Maps dengan menggunakan Google Maps *Application Program Interface* (API).



Gambar 3.1 Aktivitas Utama Proyek

### 3.3 Manfaat Bahasa Inggris

Menurut English Domain - sebuah institusi non-formal yang mengajarkan bahasa Inggris di Eropa, terdapat beberapa alasan untuk mempelajari bahasa Inggris:

- a. Berkomunikasi dengan orang lain. Bahasa Inggris merupakan bahasa yang paling umum dan sering digunakan di seluruh dunia untuk berkomunikasi antar bangsa dan negara, sehingga memudahkan setiap warga negara untuk berpergian dan berinteraksi ke negara asing.

- b. Mendorong karir jauh ke depan. Dapat berbicara bahasa Inggris berarti terbukanya kesempatan untuk berkomunikasi dengan klien dan rekan asing sehingga meningkatkan nilai tambah di dalam dunia profesional.
- c. Mendapatkan akses pengetahuan. Istilah dalam berbagai bidang pengetahuan dan akademis menggunakan bahasa Inggris sebagai bahasa pengantar sehingga membantu dalam pengumpulan informasi dan ilmu pengetahuan.
- d. Melihat dunia dengan perspektif berbeda. Bahasa Inggris memungkinkan setiap orang untuk berkesempatan merasakan berbagai perbedaan budaya dari seluruh dunia, dan juga bahasa yang digunakan dalam berbagai macam media informasi dan hiburan.

### **3.4 Perbandingan TOEFL, TOEIC, IELTS**

TOEFL, TOEIC, dan IELTS memiliki fungsi yang sama yakni menguji kemampuan bahasa Inggris dengan spesialisasi yang berbeda (Cyssco & Rasyid, 2013):

a. TOEFL

Merupakan tes bahasa Inggris akademik yang paling umum bagi orang yang berbahasa Inggris. Persyaratan di hampir seluruh universitas dan perguruan tinggi di negara berbahasa Inggris. Lembaga pemerintahan, badan perizinan, bisnis, dan program beasiswa juga mungkin membutuhkan TOEFL. Nilai TOEFL berlaku hingga dua tahun.

b. TOEIC

*Test of English for International Communication* (TOEIC) merupakan tes bahasa Inggris yang diciptakan oleh ETS atas permintaan Menteri Perdagangan dan Industri Luar Negeri Jepang untuk mengukur kemampuan bahasa Inggris pada aktivitas bisnis berskala internasional.

c. IELTS

*International English Language Testing System* (IELTS) merupakan tes bahasa Inggris yang dikembangkan oleh *University of Cambridge ESOL Examinations* bersama dengan *British Council* dan *IDP Education Australia*. Memiliki kesamaan dengan TOEFL namun lebih fokus pada perguruan tinggi di negara persemakmuran Inggris dan sebagian negara Eropa.

Berdasarkan spesialisasi tersebut, tes yang paling cocok untuk diujikan pada mahasiswa UMN adalah TOEFL karena tes bahasa Inggris akademik yang paling umum dan dapat dipakai juga sebagai pengukur kemampuan bahasa Inggris sehari – hari dalam berbagai macam bidang dan keperluan.

### 3.5 Perbandingan Metode Data Collection

Terdapat beberapa cara untuk memperoleh data, salah satunya dengan menggunakan survei. Survei juga memiliki beberapa jenis alternatif dengan perbandingan sebagai berikut (Cobanoglu, Warde, & Moreo, 2001, p. 407):

Factor	Mail	Fax	Web-based
Coverage	High	Low	Low
Speed	Low	High	High
Return Cost	Preaddressed/Prestamped envelope	800 return fax number	No cost to the respondent
Incentives	Cash/Non-cash incentives can be included	Coupons may be included	Coupons may be included
Wrong addresses	Low	Low	High
Labor needed	High	Medium	Low
Expertise to construct	Low	Medium	High
Variable Cost/each survey*	About \$1.00	About \$ 0.50	No cost

\*: These are estimated costs for a 4-page survey that has a population in the U.S.

Gambar 3.2 Perbandingan Survei

Sumber: Cobanoglu, Warde, & Moreo, 2001, p. 407

Berdasarkan hasil perbandingan di atas, survei yang paling cocok untuk pengambilan data pada riset ini adalah Web-based karena:

- a. Proses survei tidak membutuhkan biaya baik dari pihak penyelenggara maupun responden.
- b. Hasil survei yang telah selesai dikerjakan oleh responden dapat langsung dilihat dan dinilai secara otomatis.
- c. Tidak membutuhkan tenaga kerja yang banyak karena survei disalurkan melalui media sosial.

Selain itu, terdapat berbagai metode pengumpulan data berdasarkan Center of Disease Control and Prevention (CDC), sebuah badan Departemen Kesehatan dan Layanan Masyarakat Amerika Serikat, yakni :

### **3.5.1 *Focus Group***

*Focus Group* adalah sebuah kelompok wawancara yang terdiri dari enam hingga dua belas orang yang memiliki karakteristik atau kepentingan yang sama (CDC, Data Collection Method for Program Evaluation: Focus Group, 2008). Seorang fasilitator membimbing kelompok tersebut berdasarkan kumpulan topik yang telah ditentukan. Fasilitator menciptakan lingkungan yang mendorong partisipan untuk berbagi persepsi dan sudut pandang, Metode ini mengumpulkan data kualitatif, yakni data deskriptif yang tidak dapat diukur dengan angka.

*Focus Group* digunakan pada saat:

- a. Ingin memperoleh informasi mendalam pada persepsi, wawasan, sikap, pengalaman, atau kepercayaan.
- b. Untuk mengumpulkan informasi tambahan sebagai penunjang metode pengumpulan data kuantitatif.
- c. Sebagai bagian dari metode pendekatan evaluasi gabungan. Metode ini digunakan untuk meningkatkan validitas dari penemuan evaluasi dengan menggunakan teknik pengumpulan data yang bervariasi.

Kelebihan dari *Focus Group* adalah:

- a. Cepat dan mudah diatur.
- b. Kedinamisan kelompok dapat menyediakan informasi berguna yang tidak disediakan saat pengumpulan data secara individual.
- c. Berguna untuk memperoleh wawasan pada topik yang lebih sulit untuk dikumpulkan melalui metode pengumpulan data lain.

Kelemahan dari *Focus Group* adalah:

- a. Rentan terhadap prasangka fasilitator.
- b. Diskusi dapat didominasi atau berubah haluan oleh beberapa individu.
- c. Analisis data membutuhkan waktu dan perlu direncanakan secara matang.
- d. Tidak menyediakan informasi valid pada level individual.
- e. Informasi bukan perwakilan kelompok diskusi lain.

### 3.5.2 *Questionnaire*

*Questionnaire* merupakan kumpulan pertanyaan untuk mengumpulkan informasi dari individu (CDC, Data Collection Methods for Program Evaluation: Questionnaires, 2008). Pertanyaan ini dapat disebarkan melalui surat, telepon, wawancara langsung, sebagai selebaran, atau media elektronik (surat elektronik atau situs web).

*Questionnaire* digunakan pada saat:

- a. Ketika sumber daya terbatas dan membutuhkan data dari banyak orang,
- b. Untuk mengumpulkan data mengenai pengetahuan, kepercayaan, sikap, dan perilaku.
- c. Ketika perlindungan privasi responden menjadi prioritas.

Menurut Kelebihan dari *Questionnaire* adalah (Rachman, 2010) :

- a. Biasanya lebih menghemat biaya dibandingkan metode lain.
- b. Dapat mencakup wilayah yang luas.
- c. Proses pelaksanaan lebih cepat karena pembagian dapat dilakukan sekaligus.

Kelemahan dari *Questionnaire* adalah:

- a. Tidak dapat menyediakan anjuran atau bantuan terhadap responden.
- b. Tingkat respon lebih rendah dibandingkan metode lain.
- c. Adanya kemungkinan isi kuisisioner dibaca terlebih dahulu oleh responden.

### **3.5.3 *Observation***

*Observation* adalah cara pengumpulan data dengan memperhatikan sikap, kejadian, atau karakteristik fisik secara alami (CDC, Data Collection Methods for Program Evaluation: Observation, 2008). Observasi dapat dilakukan secara *overt* (diketahui oleh obyek observasi) atau *covert* (tidak diketahui oleh obyek observasi).

*Observation* digunakan pada saat:

- a. Ketika sedang memahami proses atau situasi yang sedang terjadi.
- b. Ketika mengumpulkan data pada perilaku individual atau interaksi antar manusia.
- c. Ketika pengumpulan data dari individual bukan pilihan realistik.

Kelebihan dari *Observation* adalah :

- a. Mengumpulkan data di mana dan kapan sebuah kejadian atau aktivitas terjadi.
- b. Tidak bergantung pada ketersediaan atau kemampuan orang untuk menyediakan informasi.

- c. Memungkinkan untuk melihat secara langsung apa yang orang lakukan dibandingkan bergantung pada apa yang orang katakan.

Kelemahan dari *Observation* adalah:

- a. Rentan terhadap prasangka pengamat.
- b. Rentan terhadap *hawthorne effect*, yakni orang yang bertindak lebih baik ketika mengetahui mereka sedang diawasi secara *overt*.
- c. Biaya lebih mahal dan memakan waktu dibandingkan metode pengumpulan data lain.
- d. Tidak meningkatkan pemahaman dari bagaimana orang biasanya bersikap.

#### **3.5.4 Interview**

*Interview* adalah metode yang menanyakan pertanyaan kuantitatif atau kualitatif secara oral untuk partisipan kunci (CDC, Data Collection Methods for Program Evaluation: Interview, 2009). Pertanyaan kuantitatif merupakan pertanyaan tertutup, yakni pertanyaan yang memiliki jawaban spesifik yang dipilih untuk kemudian dikategorikan dan dianalisa secara numerik. Pertanyaan kualitatif merupakan pertanyaan terbuka, yakni responden menyediakan jawaban sendiri.

*Interview* digunakan pada saat:

- a. Untuk mendapatkan informasi mendalam mengenai persepsi, wawasan, sikap, pengalaman, atau kepercayaan.
- b. Ketika mengevaluasi perbedaan individual antara pengalaman, responden, dan hasil.
- c. Sebagai penindaklanjutan metode lain.

Kelebihan dari *Interview* adalah:

- a. Berguna untuk memperoleh wawasan dan konteks menjadi topik.

- b. Memungkinkan responden untuk mendeskripsikan apa yang penting untuk mereka.
- c. Berguna untuk mengumpulkan cerita dan kutipan.

Kelemahan dari Interview adalah:

- a. Rentan pada prasangka dari pewawancara.
- b. Biaya lebih mahal dan memakan waktu dibandingkan metode pengumpulan data lain.
- c. Terkadang proses mengganggu responden.

### **3.5.5 Document Review**

*Document Review* merupakan cara mengumpulkan data dengan meninjau dokumen yang telah ada (CDC, Data Collection Methods for Evaluation: Document Review, 2009). Dokumen dapat berupa media cetak atau elektronik seperti laporan, log program, peringkat kinerja, proposal pendanaan, waktu pertemuan, dan materi pemasaran.

*Document Review* digunakan pada saat:

- a. Untuk mengumpulkan informasi latar belakang.
- b. Untuk menentukan apakah implementasi dari program mencerminkan perencanaan program.
- c. Ketika membutuhkan informasi untuk membantu mengembangkan alat pengumpulan data untuk evaluasi.
- d. Ketika membutuhkan data untuk menjawab pertanyaan evaluasi dasar yang berhubungan dengan angka.

Kelebihan dari *Document Review* adalah:

- a. Umumnya murah.
- b. Sumber yang baik untuk latar belakang.
- c. Tidak mencolok.

- d. Menyediakan pandangan dibalik layar pada program yang mungkin tidak dapat diobservasi secara langsung.
- e. Mungkin memperoleh hal yang tidak dicatat oleh orang lain.

Kelemahan dari *Document Review* adalah:

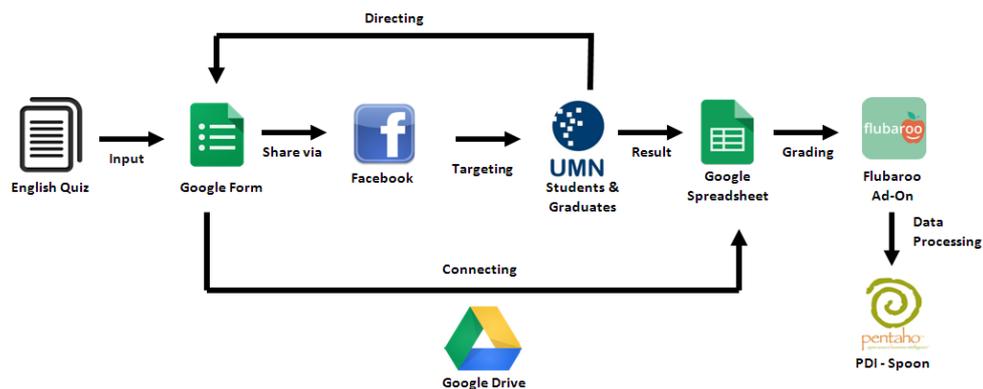
- a. Informasi mungkin tidak dapat diaplikasi, tidak tersusun, tidak tersedia, atau sudah kadaluarsa.
- b. Dapat menciptakan prasangka karena pemilihan informasi yang bertahan.
- c. Informasi mungkin tidak lengkap atau tidak akurat.
- d. Dapat memakan banyak waktu untuk mengumpulkan, meninjau, dan menganalisa banyak dokumen.

Berdasarkan hasil perbandingan dari metode pengumpulan data, maka dapat diketahui bahwa metode yang sesuai untuk riset ini adalah *Questionnaire* karena:

- a. Tidak membutuhkan sumber daya yang besar untuk mengumpulkan data.
- b. Responden tidak perlu bertatap muka secara langsung untuk melaksanakan proses pengumpulan data.
- c. Pertanyaan pada kuis dapat dikerjakan oleh beberapa responden sekaligus di lokasi yang berbeda.

### **3.6. Cara Kerja Quiz online**

Berikut ini adalah kerangka kerja kuis online yang digunakan untuk mengumpulkan hasil jawaban dari mahasiswa UMN:



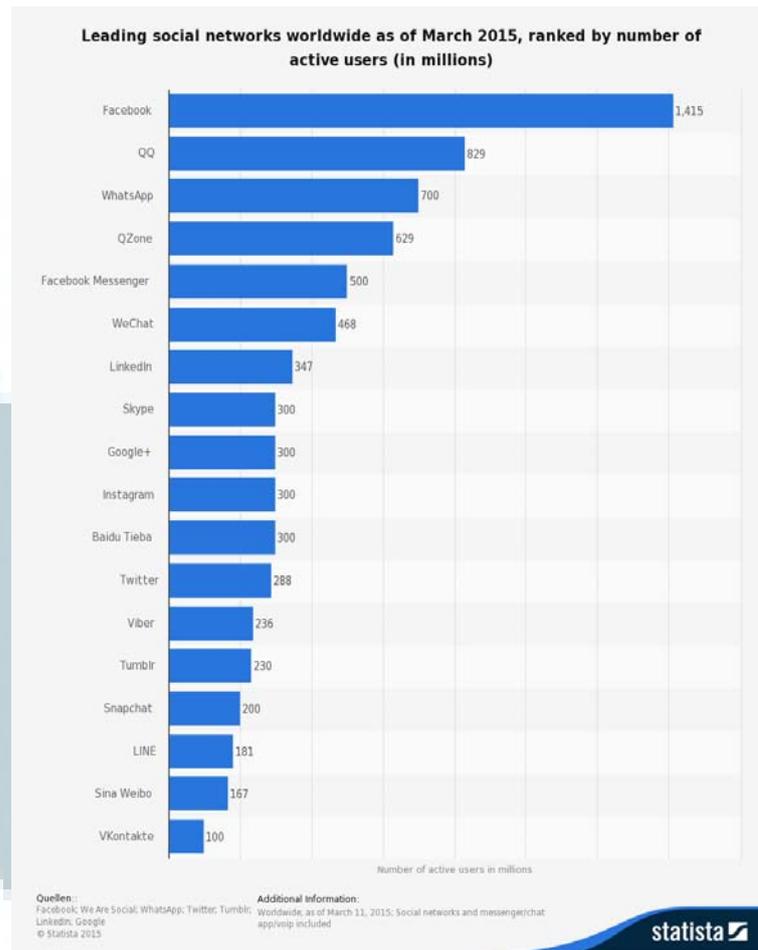
**Gambar 3.3 Kerangka Kerja Quiz online**

Kuis online yang dipergunakan di dalam proyek ini dibuat untuk mengetahui tingkat kemampuan berbahasa Inggris pada mahasiswa UMN, dengan cara membuat kuis menggunakan Google Form dan menyebarkannya di *Facebook*. Pertanyaan pada kuis tersebut berbasis pada buku latihan TOEFL (Sharpe, 2011, pp. 484-487,492-493).

Data hasil kuis tersebut kemudian dikumpulkan, dikelola, dianalisis, dan ditampilkan berdasarkan daerah yang terdapat di Indonesia untuk menemukan area mana saja yang berpartisipasi dan jumlah rata – rata dari setiap daerah. Kesenjangan dapat ditemukan melalui daerah yang kurang berpartisipasi atau hasilnya masih di bawah total rata – rata.

### **3.7 Perbandingan Sosial Media**

Berikut ini adalah hasil pengumpulan data dari sumber informasi resmi dan perhitungan statistik yang dilakukan oleh tim statista; sebuah portal statistik online, mengenai urutan sosial media berdasarkan jumlah pengguna aktif di seluruh dunia pada Maret 2015 (Statista, Leading social networks worldwide as of March 2015, ranked by number of active users (in millions), 2015):

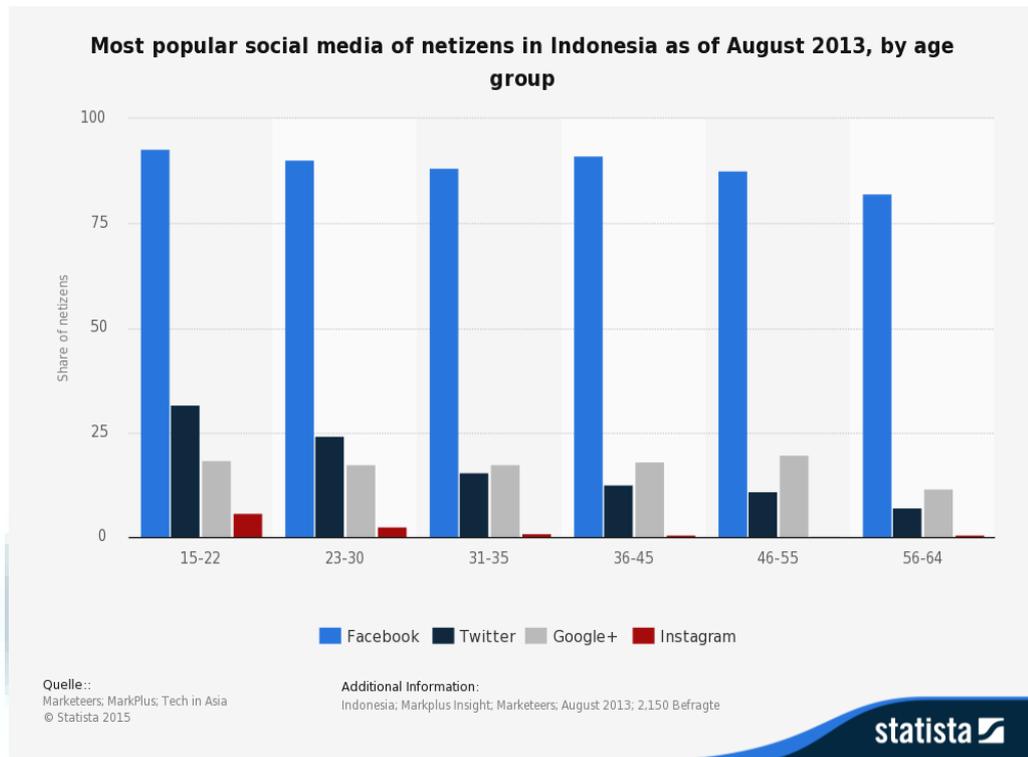


**Gambar 3.4 Perbandingan Jumlah Pengguna Aktif Sosial Media di Seluruh Dunia pada Maret 2015**

**Sumber: <http://www.statista.com>**

*Facebook* menempati urutan teratas dengan jumlah pengguna aktif mencapai lebih dari satu miliar (1.415.000.000), tidak termasuk pengguna aplikasi *Facebook Messenger* yang berada di posisi kelima. *Skype*, *Google+*, *Instagram*, dan *Baidu Tieba* memiliki jumlah pengguna aktif yang sama yakni tiga ratus juta.

Selain perbandingan pengguna media sosial di seluruh dunia, terdapat perbandingan media sosial di Indonesia pada Agustus 2013 berdasarkan kelompok usia (Statista, Most popular social media of netizens in Indonesia as of August 2013, by age group , 2013):



**Gambar 3.5 Perbandingan Media Sosial di Indonesia Tahun 2013 Berdasarkan Kelompok Usia**

Sumber: <http://www.statista.com>

Dengan total responden berjumlah 2150 orang yang tinggal di 10 kota besar di Indonesia, Facebook menempati urutan pertama pada seluruh kelompok usia dengan rata – rata jumlah pengguna diatas 75 orang, berbeda dengan media sosial lainnya yang memiliki rata – rata pengguna dibawah 50 orang.

Berdasarkan dua hasil statisik yang ada, maka dapat diketahui bahwa media sosial yang paling tepat untuk menyebarkan kuis online pada riset ini adalah *Facebook*.

### **3.8 Perbandingan Survei Online**

Alat survei online diperlukan untuk melakukan pengumpulan data kuis bahasa Inggris, berikut ini adalah empat perbandingan dari alat survei online yang dapat dipergunakan dalam kuis tersebut (Whinot, 2012):

	 SurveyMonkey	 Zoomerang	 Google Forms	 surveygizmo
Unlimited # of Surveys	✓	✗	✓	✓
Response Limits	100/survey	100/survey	200,000	1,000/month
Question Limits / Survey	10	12	255	Unlimited
Customize Brand Image	✗	✗	✓	✗
Customize Colors & Fonts	✓	✓	✓	✓
Export to PDF/ Word Doc	✗	✗	✓	✓
Embed Survey into Webpage	✓	✓	✓	✓
Randomizing and Skip Logic	Randomize (answers only)	Randomize (answers only)	✓	Randomize (answers only)
Survey Templates	✓	✓	✓	✓
Real Time Reporting	✓	✓	✓	✓
# of Question Types	15	15	6	23
Supports all Languages	✓	✓	✓	✓
Embed Images & Media	✓	✓	✓	✓
Upload Images	✓	✓	✓	✗
Support via e-mail & FAQ Tutorials	✓	✓	✓	✓
Send E-mail Using Survey Tool	✓	✓	✓	✓
Facebook Integration	✗	✓	✓	✓
Twitter Integration	✗	✓	✓	✓

Gambar 3.6 Perbandingan Survey Online

Sumber: [http:// whinot.com](http://whinot.com)

Dari beberapa kriteria yang telah disebutkan pada perbandingan di atas terdapat tiga prioritas utama yakni:

a. *Facebook integration*

Integrasi dengan Facebook diperlukan untuk mengetahui jumlah pengguna yang berpartisipasi di dalam kuis ini. Survey Monkey tidak memiliki fitur ini sehingga tidak sesuai untuk kuis ini.

*b. Upload images*

Pada kuis ini terdapat sisipan gambar pada bagian *reading* untuk menampilkan bacaan yang dibutuhkan dalam menjawab pertanyaan. Survey Gizmo tidak memiliki fitur untuk memasukkan gambar ke dalam kuis sehingga kurang sesuai untuk kuis ini.

*c. Export to PDF/ Word Doc / Excel Spreadsheet*

Data hasil kuis kemudian perlu disimpan dalam format Excel untuk digunakan pada PDI Spoon. Survey Monkey dan Zoomerang tidak memiliki fitur ini sehingga tidak cocok untuk kuis ini.

Berdasarkan hasil perbandingan dan prioritas, maka dapat diketahui bahwa Google Form memenuhi seluruh kriteria yang telah ditentukan prioritasnya sehingga cocok sebagai survei online dalam kuis.

### **3.9 Fungsi Data Warehouse**

DW memiliki peranan yang sangat besar bagi penggunanya, yakni sebagai berikut (BI-Insider, 2011):

a. Peningkatan pada BI :

Wawasan dapat diperoleh melalui akses informasi yang ditingkatkan. Pengambil keputusan tidak akan dibebaskan dalam mengambil keputusan berdasarkan firasat semata. Keputusan yang mempengaruhi strategi dan operasi dari organisasi berdasarkan fakta yang masuk akal dan disokong menggunakan bukti dan data aktual organisasi. Terlebih lagi, pembuat keputusan akan diinformasikan dengan lebih baik saat mereka mampu mengambil data aktual dan memperoleh informasi berdasarkan kebutuhan personal mereka.

b. Meningkatkan performa dari query dan sistem :

DW dirancang dan dikonstruksikan dengan fokus pada kecepatan dalam mengambil data dan analisi, dan dirancang untuk menyimpan data dalam jumlah besar mampu mengambil data secara cepat. Sistem analisis dikonstruksikan secara berbeda dari sistem operasional yang fokus pada pembuatan dan modifikasi data. Kemudian, DW dibangun untuk analisa dan memperoleh data dibandingkan perawatan efisien dari laporan individual.

c. Memungkinkan memperoleh BI dari beberapa sumber :

Untuk beberapa organisasi, sistem informasi terdiri dari beberapa sub-sistem yang terpisah dan tercipta dari platform yang berbeda. Terlebih lagi, menggabungkan data dari beberapa sumber yang terpisah merupakan kebutuhan yang umum dalam menjalankan BI. Untuk menyelesaikan permasalahan ini, DW melakukan integrasi dari data yang terpisah dan membuatnya dapat diakses melalui satu tempat.

d. Mengakses data tepat waktu :

DW memungkinkan pengguna dan pembuat keputusan untuk mengakses data dari berbagai sumber yang diperlukan. Pengguna akan membutuhkan waktu yang sedikit untuk proses pengambilan data. Jadwal rutin integrasi data, yang dikenal sebagai ETL, dipengaruhi oleh lingkungan DW. Rutinitas ini mengumpulkan data dari beberapa sumber dan mengubah data menjadi bentuk yang berguna. Kemudian pengguna dapat dengan mudah mengakses data dari satu tampilan dan dapat mengambil data secara langsung dengan dukungan TI yang lebih sedikit. Hal ini memungkinkan pengguna untuk menggunakan waktu lebih banyak melakukan analisa data dan lebih sedikit pengumpulan data.

e. Meningkatkan kualitas dan konsistensi data :

Implementasi DW umumnya termasuk konversi data dari beberapa sumber sistem, file data, dan perubahan dari data yang berbeda menjadi bentuk umum. Data tersebut distandarisasi dan yang tidak konsisten dihilangkan. Data pada departemen dan unit dari organisasi akan menggunakan penyimpanan data yang sama sebagai sumber individual untuk laporan dan pengambilan data.

f. **Intelijensi historis**

Keuntungan dari DW adalah memungkinkan pelaporan dan juga dapat menggunakan analisis dengan menggunakan periode waktu, analisa trend, dan juga melakukan prediksi berdasarkan hasil analisis sebelumnya menggunakan sumber data yang beragam.

g. **Return on Investment (ROI) yang tinggi**

ROI adalah jumlah keuntungan yang diterima atau penurunan biaya untuk merealisasikan proyek dan investasi, kemudian melakukan alokasi dana yang berlebih secara lebih efektif.

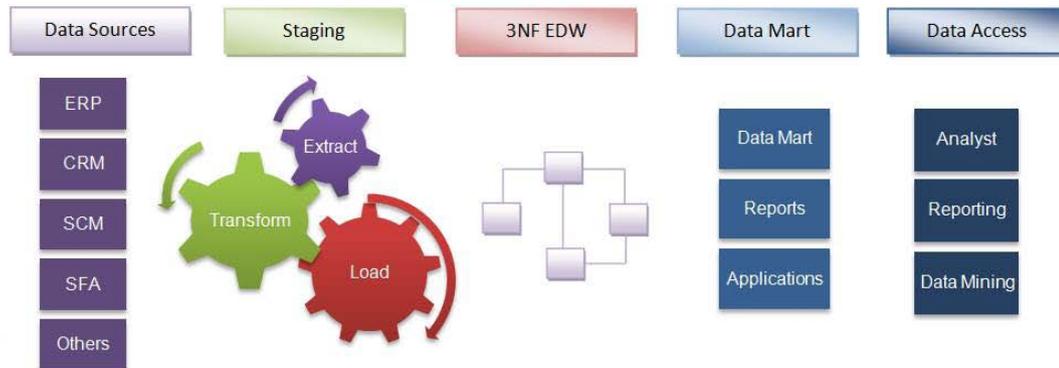
### **3.10 Metodologi Data Warehouse**

Di dalam pembuatan DW, diperlukan adanya suatu metodologi yang sesuai dengan kebutuhan dari perusahaan atau organisasi sehingga mengoptimalkan penggunaannya. Untuk pendekatan terdapat Data Mart berdasarkan model Ralph Kimball dan pendekatan EDW berdasarkan model Bill Inmon.

a. **Pendekatan *Top-Down* Bill Inmon**

Menurut George dalam Search Business Intelligence.in, Inmon mendefinsikan data warehouse sebagai penyimpanan yang tersentralisasi untuk keseluruhan perusahaan. Data warehouse menyimpan data ‘atomik’ ke detil dengan level terkecil. *Data mart* dimensional diciptakan hanya setelah data warehouse lengkap telah diselesaikan, jadi data warehouse berada di

tengah *Corporate Information Factory* (CIF), yang menyediakan *framework* logikal untuk menyampaikan BI.

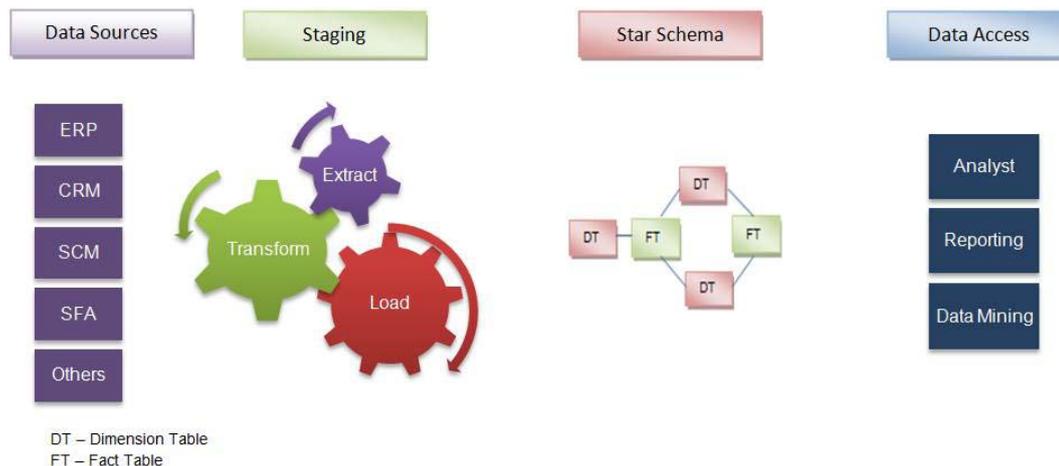


Gambar 3.7 Diagram Top-Down Bill Inmon

Sumber: <http://searchbusinessintelligence.techtarget.in>. Akses tanggal 29/03/2015

b. Pendekatan *Bottom-Up* Ralph Kimball :

Kimball mendefinisikan DW sebagai salinan data transaksi yang distruktur secara khusus untuk query dan analisis. Arsitektur DW Kimball juga diketahui sebagai Bus DW, yakni jalur transmisi pada sinyal yang diturunkan dan diambil pada setiap alat yang terhubung ke saluran tersebut, dan hanya alat yang menjadi tujuan yang akan menangkap sinyal tersebut, selebihnya menghilangkan sinyal tersebut. Menurut Winn L. Rosch, dapat dianalogikan sebagai bus yang berhenti pada setiap halte untuk menurunkan dan menaikkan penumpang. Model dimensional fokus pada kemudahan bagi akses pengguna akhir dan menyediakan kinerja tingkat tinggi untuk DW.



**Gambar 3.8 Diagram Bottom-Up Ralph Kimball**

Sumber: <http://searchbusinessintelligence.techtarget.in>. Akses tanggal 29/03/2015

Berikut ini adalah tabel mengenai perbandingan antara pendekatan Data Mart dengan EDW:

**Tabel 3.1 Perbandingan Data Mart dan EDW**

Sumber: Turban, Sharda, Delen, King, & Aronson, 2011, p. 53

Faktor Perbandingan	Data Mart	EDW
Cakupan	1 Wilayah Subyek	Beberapa Wilayah Subyek
Waktu Pengembangan	Beberapa Bulan	Beberapa Tahun
Biaya Pengembangan	\$10,000 - \$100,000+	\$1,000,000
Kesulitan Pengembangan	Rendah – Sedang	Tinggi
Syarat Pembagian Data	Umum (di dalam area bisnis)	Umum (seluruh perusahaan)
Sumber	Hanya beberapa sistem operasional dan eksternal	Banyak sistem operasional dan eksternal
Ukuran	MB sampai beberapa GB	GB sampai PB (Petabyte)
Rentang Waktu	Mendekati saat ini & Data	Data historis

	historis	
Transformasi Data	Rendah – Sedang	Tinggi
Frekuensi Update	Jam-an, Harian, Mingguan	Bulanan, Tahunan
Teknologi Hardware	Workstation & Server Departemen	Server Perusahaan & Komputer Mainframe
Sistem Operasi	Windows & Linux	Unix, Z/OS, OS/390
Database	Workgroup / Server DB Standar	Server Database Enterprise
Jumlah Pengguna Serentak	10an	100an - 1000an
Jenis Pengguna	Business Area Analyst & Manager	Enterprise Analyst & Senior Executive
Perhatian Bisnis	Optimasi aktivitas dalam area bisnis	Optimasi fungsional silang & pengambilan keputusan

Berdasarkan tabel diatas, dapat diketahui bahwa pendekatan *data mart* lebih sesuai untuk proyek ini dibandingkan dengan EDW, dengan alasan:

- a. Proyek ini hanya memiliki satu wilayah subyek, yakni pendidikan.
- b. Proyek ini membutuhkan waktu pengembangan yang cepat, sehingga tidak mungkin memilih EDW yang memakan waktu pengembangan sampai bertahun – tahun.
- c. Frekuensi pembaharuan berlangsung per hari.
- d. Proyek ini tidak membutuhkan ukuran data yang sangat besar.

Kemudian terdapat tabel perbandingan antara model Bill Inmon dan Ralph Kimball (Turban, Sharda, Delen, King, & Aronson, 2011):

**Tabel 3.2 Perbandingan Inmon dan Kimball**

Sumber: Turban, Sharda, Delen, King, & Aronson, 2011, p. 54

	Karakteristik	Inmon	Kimball
Metodologi & Arsitektur	Keseluruhan pendekatan	Top-Down	Bottom-Up
	Susunan Arsitektur	DW Enterprise-Wide (Atomic) 'feed' DB departemen	Model data mart proses bisnis tunggal & konsistensi perusahaan diperoleh melalui data bus & dimensi yg sesuai
	Kerumitan Metode	Cukup rumit	Cukup sederhana
	Perbandingan dengan metodologi pengembangan yang dihasilkan	Berasal dari metodologi spiral	4 Langkah Proses: dari metode DBMS
	Diskusi dari rancangan fisik	Cukup Menyeluruh	Cukup Ringan
Data Modelling	Orientasi Data	Berbasis Subyek / Data	Orientasi Proses
	Alat	Tradisional (ERD, data flow diagram)	Model Dimensional; dari model relasional
	Aksesibilitas Pengguna	Rendah	Tinggi
Philosophy	Pengguna Utama	Profesional IT	Pengguna Akhir
	Tempat di dalam organisasi	Bagian tak terpisahkan dari pabrik informasi perusahaan	Transformer & Retainer dari data operasional
	Tujuan	Menyediakan solusi teknis berbasis metode DB & teknologi yang terbukti	Menyediakan solusi yang memudahkan pengguna akhir u/ query data secara langsung & dapat waktu respon yang cepat

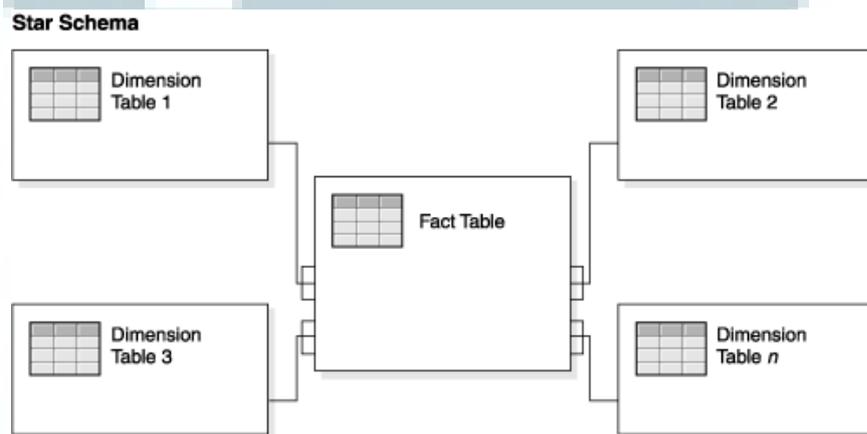
Berdasarkan tabel diatas, model yang sesuai untuk proyek ini adalah Kimball karena :

- a. Pendekatan proyek ini dilakukan dari hasil akhir hingga pengumpulan kebutuhan.
- b. Metode yang digunakan cukup mudah.
- c. Alat yang digunakan adalah model dimensional.
- d. Fokus pada pengguna akhir.

### 3.11 Data Model

Terdapat berbagai jenis struktur data model dimensional yang diciptakan oleh Ralph Kimball, namun hanya tiga skema yang sering digunakan yakni:

#### 3.10.1. Star Schema



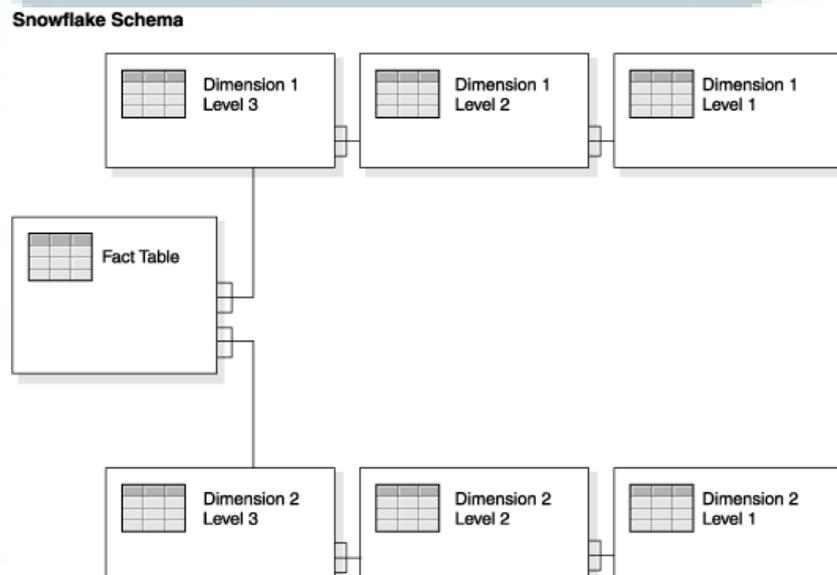
Gambar 3.9 Bagan Star Schema

Sumber: <http://www-01.ibm.com>

1. Keunggulan:
  - a. Struktur data sederhana sehingga lebih mudah dipahami.
  - b. Bentuk skema yang paling umum, mudah terintegrasi dengan alat lain.

- c. Pemrosesan query lebih mudah karena tidak ada proses normalisasi.
  - d. Meningkatkan kinerja karena denormalisasi.
  - e. Optimal untuk data besar karena kecepatan kinerja sistem dan query.
  - f. Aksi agregasi yang cepat.
  - g. Baik untuk OLAP.
2. Kelemahan:
- a. Integritas data kurang, terdapat duplikasi data dan data anomali.
  - b. Pemuatan tabel dimensi yang memakan waktu lebih lama.
  - c. Membutuhkan lebih banyak ruang pada DB.
  - d. Sulit membuat query yang lebih rumit.
  - e. Skema tersebut tidak dapat melakukan relasi 'many-to-many'.

### 3.10.2. Snowflake Schema



**Gambar 3.10 Bagan Snowflake Schema**

Sumber: <http://www-01.ibm.com>

1. Keunggulan:

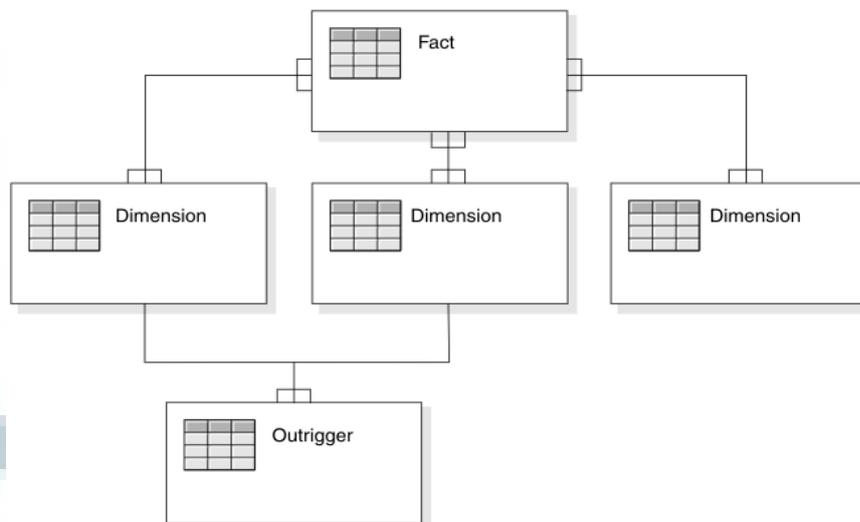
- a. Kualitas data lebih baik; tidak ada duplikasi maupun anomali.
- b. Lebih hemat ruang pada DB.
- c. Kinerja query lebih spesifik, optimasi untuk tampilan yang spesifik.
- d. Terdapat beberapa alat yang dirancang khusus untuk data khusus di organisasi.
- e. Lebih banyak data yang terstruktur.

2. Kelemahan:

- a. Struktur ini memaksakan normalisasi pada tabel dimensi dan tugas proses data ditambahkan untuk menjamin kualitas data.
- b. Struktur lebih rumit dibandingkan skema bintang.
- c. Kinerja query kurang karena banyaknya operasi join untuk mengambil data.

### 3.10.3. Starflake Schema

Starflake Schema



Gambar 3.0.11 Bagan Starflake Schema

Sumber: <http://www-01.ibm.com>

1. Keunggulan:

- a. Mengurangi *join* dan pengoperasian query lebih cepat.
- b. Penulisan query menjadi lebih mudah karena tidak ada join yang rumit.
- c. Pekerjaan pengguna akhir jadi lebih mudah karena kerumitan berkurang.

2. Kelemahan:

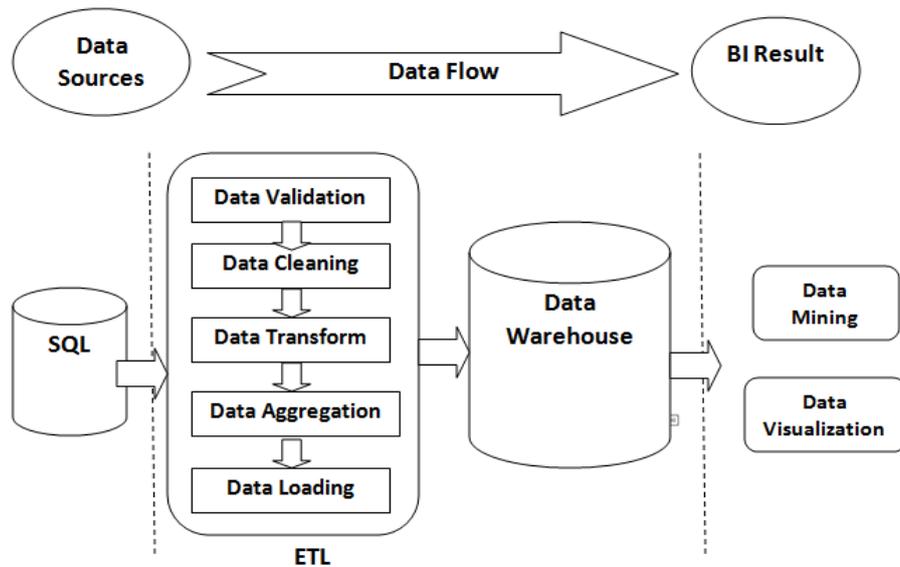
- a. Ukuran tabel menjadi lebih besar karena banyaknya jumlah tabel dimensi.
- b. Versi denormalisasi dari skema kepingan salju.
- c. Terlalu banyak baris di tabel fakta.

Berdasarkan hasil perbandingan tiga skema diatas, riset ini lebih optimal menggunakan skema bintang karena:

- a. Skema ini merupakan skema yang paling sederhana dan juga yang paling sering digunakan sehingga mudah diaplikasikan di segala platform.
- b. Mudah dipahami oleh anggota riset lainnya sehingga mengurangi tingkat kesalahan atau kebingungan.
- c. Data pada riset ini tidak banyak dan berat sehingga tidak diperlukan proses normalisasi untuk mengurangi penggunaan kapasitas sistem.

### **3.12 Tahapan Proses Data Warehouse**

Terdapat beberapa tahapan di dalam proses DW, yang dijelaskan melalui gambar di bawah ini:



Gambar 3.12 Data Warehouse Workflow

Sumber: <http://www.intechopen.com>. Akses tanggal 12/03/2015

Proses *data warehouse* yang pertama adalah mengambil data dari sumber data yang sudah ditetapkan, pada proyek ini yakni hasil proses pengumpulan data dalam bentuk *Comma Separated Values* (CSV) atau Excel, kemudian dilanjutkan dengan proses ETL dengan PDI Kettle. Validasi data (*data validation*) dilakukan dengan menggunakan step *Data Validator* untuk memastikan format data yang masuk sudah sesuai dengan standar yang telah ditentukan. Kemudian dilanjutkan dengan pembersihan data (*data cleaning*) dengan cara membentuk tabel – tabel dimensi yang telah ditetapkan sebelumnya untuk proyek ini.

Tahapan selanjutnya adalah pengumpulan data (*data aggregation*) pada tabel dimensi yang terpisah dengan cara menggabungkan data dimensi tersebut ke dalam tabel fakta dan yang terakhir adalah pemasukan data (*data loading*) ke *data warehouse* dalam bentuk file CSV atau tipe data lain yang disesuaikan untuk keperluan penambahan data dan juga visualisasi data.

### 3.13 Metode Perancangan Sistem

Perancangan DW menggunakan *Nine-Step Methodology* yang dikembangkan oleh Ralph Kimball (Kimball, *Letting the User Sleep*, 1996). Metodologi tersebut menggunakan tahapan yang dibutuhkan untuk mendesain DM, yang kemudian diintegrasikan untuk membuat DM:

#### 3.10.1 *Choose the Process*

Proses atau fungsi merupakan hal penting bagi basis data. Hal tersebut merujuk pada subyek yang perlu diperhatikan pada *datamart*. *Datamart* (DM) merupakan elemen atau komponen dari DW. Menurut Kimball, DM adalah kumpulan dari area subyek yang diatur untuk mendukung keputusan berdasarkan kebutuhan bagi setiap departemen. Sementara itu Bill Inmon berpendapat bahwa DM adalah unit dari DW yang fokus pada subyek permasalahan tertentu dari seluruh perusahaan, seperti DM untuk *finance* dan *marketing*.

Faktor yang menentukan sukses tidaknya DM adalah tingkat ketepatan waktu, kesesuaian dengan biaya yang telah ditetapkan, dan dapat memberikan solusi bagi berbagai jenis permasalahan dalam bisnis.

#### 3.10.2 *Choosing the Grain*

Yang dimaksud dengan '*choosing the grain*' adalah menentukan data apa saja yang perlu ditampilkan pada tabel fakta. Tabel fakta (*fact table*) merupakan tabel inti yang berfungsi sebagai representasi dari model basis data dan memiliki *foreign key* yang memiliki fungsi sebagai penghubung bagi *primary key* dari setiap tabel dimensi.

#### 3.10.3 *Identifying and Conforming the Dimensions*

Tabel dimensi digunakan untuk memperoleh informasi yang lebih mendetil pada DM. Tabel dimensi merupakan penggerak dari DM, platform untuk

menelusuri batasan nilai yang diperbolehkan pada DM dan kemudian memberlakukannya. Tabel dimensi merupakan sumber dari baris judul laporan pengguna akhir; mereka membawa pembendaharaan kata perusahaan ke pengguna. Kumpulan tabel dimensi yang dibangun dengan baik membuat DM menjadi lebih mudah dimengerti dan digunakan. Kumpulan tabel dimensi yang kurang ditampilkan dengan baik atau kurang lengkap akan menghilangkan kegunaan dari DM. (Ralph Kimball, 1996)

#### **3.10.4 Identify Facts**

Identifikasi fakta yang akan digunakan pada DM ditentukan melalui *grain* untuk table fakta yang telah dilakukan sebelumnya. Seluruh fakta memiliki tingkat yang sama satu dengan yang lainnya. Fakta lain dapat ditambahkan selama konsisten terhadap fakta *grain*.

#### **3.10.5 Storing pre-calculations in the Fact Table**

Fakta yang telah diidentifikasi kemudian diperiksa kembali untuk menentukan apakah fakta tersebut memerlukan perhitungan awal. Menyimpan hasil perhitungan awal dilakukan dalam pembuatan DM untuk kasus yang berhubungan dengan laba rugi.

#### **3.10.6 Rounding out the Dimensions Tables**

Pada tahap ini tabel fakta dan tabel dimensi sudah dimengerti sepenuhnya. teks keterangan ditambahkan pada table dimensi untuk menjelaskan atribut dan deskripsi lain. Teks jangan menggunakan bahasa sandi atau kode yang dapat membingungkan pengguna.

#### **3.10.7 Choosing the Duration of the Database**

Durasi mengukur seberapa lama cakupan waktu yang digunakan pada tabel fakta. Lama durasi bervariasi tergantung bagaimana perusahaan atau organisasi

menggunakan DW atau DM tersebut. Semakin jauh durasi waktu pada tabel fakta, semakin sulit untuk sumber data.

### **3.10.8 Tracking Slowly Changing Dimension**

Menggunakan versi lama dari tabel dimensi yang penting, bukan versi terbaru merupakan keharusan. Hal ini dikenal sebagai permasalahan *slowly changing dimension*. Deskripsi yang sesuai mengenai produk dan pelanggan lama harus digunakan dengan histori transaksi yang lama. (Kimball, 1996)

Terkadang, kunci yang digeneralisasi harus diberikan pada tabel dimensi penting untuk membedakan beberapa cuplikan isi dari tabel dimensi selama periode waktu tertentu.

Terdapat tiga jenis *slowly changing dimensions* dasar yakni:

- a. Perubahan pada atribut dimensi dilakukan dengan penulisan ulang (*overwritten*).
- b. Perubahan pada atribut dimensi menyebabkan pembuatan catatan dimensi yang baru.
- c. Perubahan pada atribut dimensi menyebabkan pembuatan atribut alternatif sehingga atribut lama dan baru dapat diakses secara bergantian dari catatan dimensi yang sama.

### **3.10.9 Deciding the Query Priorities and the Query Modes**

Sebagian besar permasalahan desain fisik penting yang mempengaruhi persepsi pengguna akhir termasuk:

- a. Pengurutan tabel fakta secara fisik pada disk penyimpanan
- b. Keberadaan ringkasan atau agregasi sebelum disimpan dan dapat berupa alat desain yang signifikan pada DW.

Permasalahan desain fisik yang lain termasuk administrasi, cadangan (*backup*), kinerja penindeksan, dan keamanan.

### **3.14 Platform Data Warehouse**

Selain memilih metodologi untuk *data warehouse*, diperlukan platform yang sesuai untuk menjalankan proses DW. Sebelum menentukan platform terlebih dahulu memilih antara sumber terbuka (*open source*) dengan komersial.

Kelebihan dari sumber terbuka adalah (Cloud Tweaks, 2011):

- a. Pengembangan lebih dapat disesuaikan karena kode dapat diakses oleh siapa saja.
- b. Dapat dengan mudah memperbaiki gangguan tanpa harus menunggu perbaikan dari pihak pengelola.
- c. Tidak bergantung pada pihak pengembang atau pengelola.
- d. Menggunakan standar umum dapat menerima hampir seluruh bentuk file atau data.

Sementara itu, kelemahan dari penggunaan sumber terbuka adalah:

- a. Tidak dapat langsung dipergunakan, harus disesuaikan dengan kebutuhan terlebih dahulu.
- b. Diperlukan waktu untuk latihan sebelum dapat menguasainya.
- c. Beberapa perangkat keras terbaru tidak dapat menerima sumber terbuka.

Kelebihan dari komersial adalah (Business Bee, 2014) :

- a. Telah diuji dan digunakan oleh banyak pengguna.
- b. Integrasi umumnya lebih cepat dan mudah.
- c. Kemungkinan terdapat gangguan lebih sedikit karena dikerjakan oleh profesional.

Sementara itu, kekurangan dari komersil adalah:

- a. Membutuhkan biaya pendukung dan perbaikan yang besar.
- b. Membutuhkan waktu yang lebih untuk melakukan perbaikan.
- c. Bergantung pada pihak pengembang.

Berdasarkan pertimbangan di atas, maka menggunakan sumber terbuka lebih sesuai untuk pengembangan proyek karena:

- a. Dapat melakukan penyesuaian
- b. Tidak membutuhkan waktu yang lama dalam pengembangan.

Kemudian terdapat beberapa pilihan aplikasi sumber terbuka yang dapat digunakan dalam proyek ini, yaitu:

- a. Talend Open Source ETL Tool (Talend Open Studio)
  - i. Kelebihan:
    - (a). Tool yang mudah digunakan yang sudah berada dalam platform Java dan dengan mudah diimplementasikan ke server pengguna.
    - (b). Menawarkan lingkungan model grafis yang ramah bagi pengguna dan fungsi lain yang mungkin tidak ada di tool sumber terbuka lainnya.
    - (c). Menyediakan solusi perangkat lunak yang fleksibel menurut model bisnis yang berbeda terlepas dari ukurannya.
  - ii. Kekurangan:
    - (a). Tool ini cukup menyulitkan bagi orang yang baru di integrasi data.
    - (b). Tidak cocok untuk memproses data yang besar.
- b. Clover ETL
  - i. Kelebihan:

- (a). Menawarkan fleksibilitas dan kenyamanan bagi pengguna dengan kemudahan transformasi dan integrasi data.
  - (b). Proses yang cepat dan akurasi merupakan poin keunggulan dari tool ini.
  - (c). Merupakan keseimbangan antara penyuntingan transformasi grafis dan fungsi program.
- ii. Kekurangan:
- (a). Pilihan pengguna untuk sambungan masuk dan keluar dibatasi.
  - (b). Ketidakhadiran plug-in tambahan untuk menambah fitur.
- c. Pentaho (Kettle)
- i. Kelebihan:
- (a). Merupakan rangkaian BI yang cukup dapat diandalkan dan berkembang pesat di pasaran.
  - (b). Mudah bagi pengguna yang tidak membutuhkan keahlian dalam integrasi.
  - (c). Memiliki dukungan komunitas yang kuat.
  - (d). Pengguna menganggapnya fleksibel dalam proses ETL.
- ii. Kekurangan:
- (a). Merupakan tool debugging yang sangat dasar.
  - (b). Memiliki pilihan yang terbatas dalam menambahkan fungsi.
  - (c). Proses intepretasi lebih lambat dibandingkan ETL tool lainnya.

Berdasarkan hasil diatas, maka platform yang sesuai untuk proyek ini adalah Kettle dari Pentaho Data Integration (PDI) karena:

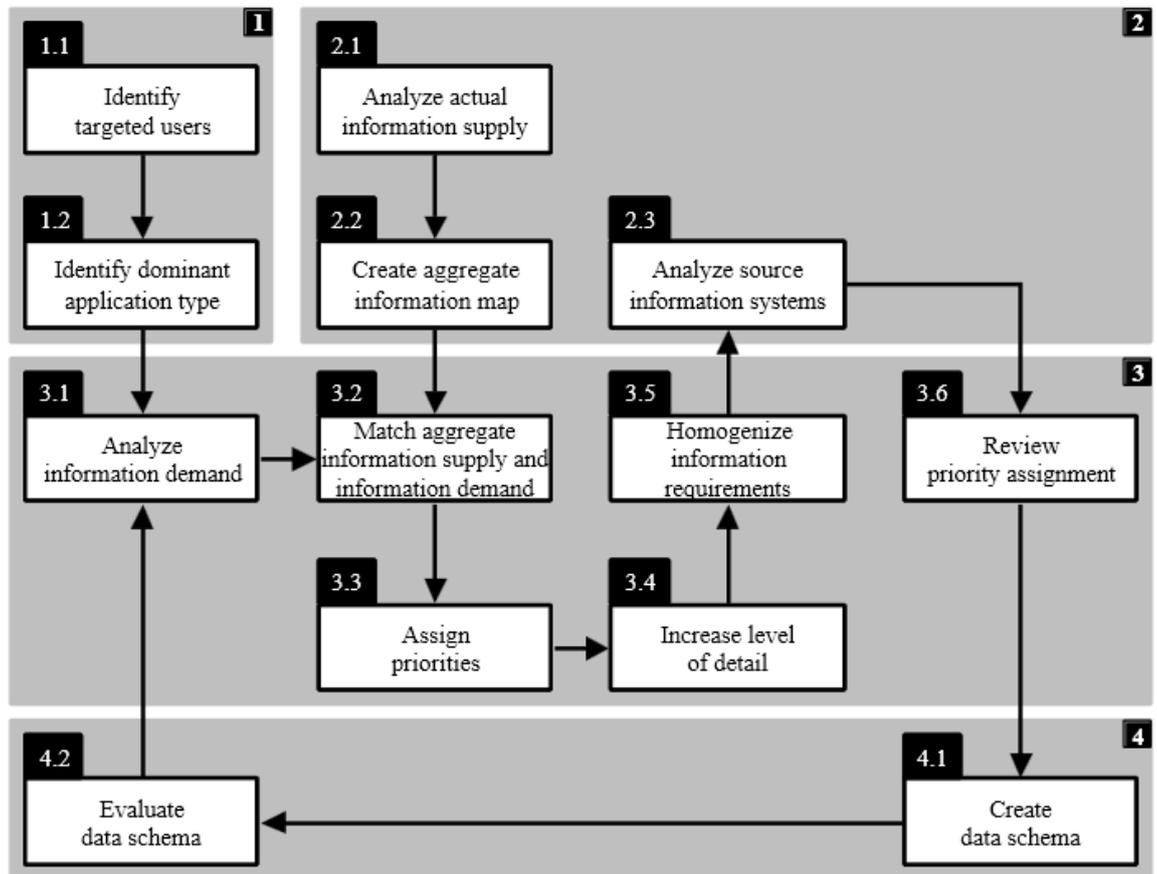
- a. Menerima masukan dan dapat mengeluarkan bentuk data yang bervariasi.
- b. Memiliki banyak sumber informasi dan juga dukungan dari komunitas.
- c. Proses ETL fleksibel.

### 3.15 *Requirement Analysis*

Untuk memastikan bahwa DM yang dibuat sesuai dengan kebutuhan dari pengguna akhir, maka diperlukan adanya analisis kebutuhan. Saat ini terdapat dua jenis analisis kebutuhan yang digunakan, namun masing – masing memiliki kelemahan yakni (Winter & Strauch, 2003):

- a. Pendekatan berbasis penawaran (*supply-driven*) memiliki resiko untuk menyalakan sumber daya dengan struktur informasi yang tidak dibutuhkan dan kurang melibatkan pengguna DW.
- b. Pendekatan berbasis permintaan (*demand-driven*) terkadang menderita karena spesifikasi dari proses manajemen bersangkutan yang terlalu abstrak, dari mekanisme pencocokan antara suplai informasi dan permintaan informasi yang tidak sesuai atau kurangnya dukungan untuk penentuan prioritas dan penyeragaman kebutuhan informasi.

Oleh sebab itu Robert Winter dan Benhart Strauch menciptakan analisis kebutuhan dengan memisahkan antara analisis kebutuhan informasi (tahap 1) dan pengembangan sistem informasi (tahap 2). Kemudian spesifikasi kebutuhan informasi terdapat pada tahap 3, dan tahap 4 merupakan hasil perbandingan antara hasil yang diharapkan dan kenyataannya. Berikut ini adalah tahapan dari pembuatan analisa kebutuhan DW hasil rancangan dari Winter dan Strauch:



Gambar 3.13 Model Aktivitas untuk Requirement Analysis

Sumber: Information Requirement Engineering for Data Warehouse Systems

(Winter & Strauch, 2003)

### 3.14.1 Tahap 1: Inisialisasi

#### a. Identifikasi sasaran pengguna

Pendekatan secara bertahap digunakan untuk pengembangan DW secara umum, setiap siklus perlu diidentifikasi sasaran pengguna atau proses keputusan dari siklus tersebut.

Contoh: Siklus pengembangan untuk DW di bank hanya dapat menargetkan laporan untuk keamanan transaksi.

## **b. Identifikasi tipe aplikasi dominan**

Tergantung pada jenis pengguna (contoh: manajemen menengah atau knowledge worker) dan jenis analisis (contoh: laporan biasa atau analisa pengecualian drill-down), jenis aplikasi analisa yang berbeda digunakan. Jenis aplikasi yang dominan harus diidentifikasi karena spesifikasi ini memiliki maksud pada data model mana yang digunakan, sumber sistem manakah yang berhubungan, dan proses keputusan yang harus diambil.

### **3.14.2 Tahap 2: Analisis 'As is'**

#### **a. Analisa suplai informasi aktual**

Sumber informasi yang paling penting dari sumber informasi aktual adalah analisa laporan yang rutin digunakan oleh pengguna yang berhubungan. Dengan menciptakan penyimpanan dari laporan ini, informasi yang dimasukkan dan model data dasar dapat dicapai. Tidak hanya kebutuhan informasi lebih mudah dicocokkan dengan informasi yang tersedia dan seleksi dari laporan yang sesuai dapat didukung, namun pembuatan metadata juga mendukung perkembangan aktivitas DW aktual dan di masa mendatang.

#### **b. Menciptakan kumpulan peta informasi**

Untuk memandu proses pencocokan dari persediaan dan permintaan informasi, diperolehlah kumpulan 'peta informasi'. Input untuk aktivitas ini adalah metadata yang diperoleh dari analisis laporan. 'peta informasi' tersebut dasarnya adalah sebuah skema data dari subyek informasi yang berhubungan pada tingkat keseluruhan.

### **c. Analisa sumber sistem informasi**

Sebuah analisa yang akurat dari sumber data yang berhubungan (contoh: transaksi sistem yang terpilih atau tampilannya) merupakan kewajiban untuk memastikan tingkat kualitas data cukup memadai. Kontrol kualitas data yang efisien diperlukan untuk memastikan keberhasilan implementasi DW dan penerimaan dari pengguna.

#### **3.14.3 Tahap 3: Analisis 'To be'**

##### **a. Analisa permintaan informasi**

Beberapa perusahaan menggunakan apa yang disebut 'pertanyaan bisnis' untuk menganalisa proses keputusan dan mengidentifikasi kebutuhan informasi yang bersangkutan. 'Pertanyaan bisnis' merupakan pertanyaan tipikal yang berhubungan dengan pengguna atau proses keputusan dan yang tidak dapat dijawab menggunakan sistem informasi yang ada saat ini. Informasi yang hilang merupakan permasalahan utama bagi pembuat keputusan atau *knowledge worker*. Sangatlah penting untuk Identifikasi 'pertanyaan yang tepat'. Sistem DW memiliki suplai informasi untuk menjawab pertanyaan bisnis tersebut.

##### **b. Mencocokkan kumpulan suplai informasi dan permintaan informasi**

Permintaan informasi untuk pertanyaan bisnis yang sesuai dicocokkan dengan kumpulan peta informasi untuk membuat analisis awal atau kasar dari tujuan kebutuhan informasi yang belum terpenuhi. Berdasarkan analisis tersebut, keputusan harus dibuat dalam aktivitas selanjutnya pada kebutuhan informasi yang harus disediakan oleh sistem DW.

### **c. Menetapkan prioritas**

Untuk menetapkan prioritas bagi kebutuhan informasi, terdapat pendekatan berbasis pada biaya pengembangan, waktu implementasi, kapasitas dari tim pengembang, aspek keamanan data, aspek privasi, rincian informasi yang dikehendaki, frekuensi refresh yang dikehendaki, aspek kualitas data, biaya kualitas data, dan kombinasi dari kriteria tersebut.

### **d. Meningkatkan tingkat detail**

Sejak penyocokan awal dari suplai dan permintaan informasi sudah dilakukan pada tingkat keseluruhan, penyempurnaan merupakan hal yang penting. Untuk kebutuhan informasi yang telah dipilih sebelumnya, sumber data dianalisa dan peraturan transformasi dispesifikasi. Tambahannya, dispesifikasi data apa saja yang disediakan di level keseluruhan atau rincian dan seberapa sering data perlu di-*refresh*.

### **e. Menyeragamkan kebutuhan informasi**

Data disediakan oleh sistem DW harus diasosiasikan dengan konsep semantik yang didefinisikan dengan jelas untuk memiliki arti dan dapat diutilitas oleh pengguna. Sayangnya arti semantik banyak data terkadang tidak seragam di seluruh organisasi besar. Tidak konsisten dan sebagai konsekuensinya harus menerima akibat dari permasalahan tersebut. Oleh karena itu arti semantik dari data harus diseragamkan dalam organisasi untuk menghindari permasalahan penerimaan dan tidak konsisten.

### **f. Meninjau penetapan prioritas**

Spesifikasi dari kebutuhan informasi yang telah diseragamkan dapat ditinjau dengan analisis akurat dari sumber data yang berhubungan.

Kebutuhan informasi dapat dinilai berdasarkan biaya suplai, biaya proses ekstraksi atau transformasi, tingkat kualitas yang diharapkan, tingkat kepraktisan pelayanan suplai, dan lainnya. Oleh karena itu membuat prioritas informasi mendetil dan seragam yang harus disediakan oleh sistem DW.

#### **3.14.4 Tahap 4: Modeling**

##### **a. Membuat skema data**

Skema data berfungsi sebagai dasar untuk tahap pengembangan sistem berikutnya dan sumber yang sangat penting untuk metadata. Untuk modeling skema data (contoh: *Entity Relationship Model*), model data yang sesuai harus dipilih terlebih dahulu. Tujuan dari data model semantik umum adalah untuk digunakan secara luas bagi model skema data sistem informasi biasa. Data model semantik spesifik data multidimensional baru mulai diterima untuk skema data modeling dari sistem berorientasi keputusan.

Meskipun model semantik untuk data multidimensional jauh dari sempurna, tidak dianjurkan untuk menggunakan model data semantik umum untuk data multidimensional karena model ini telah dikembangkan untuk menjaga konsistensi dari DB yang stabil dengan rancangan yang sesuai. Kontrasnya, DB untuk mendukung keputusan atau *knowledge work* sebagian besar tidak stabil dan pengulangan tertentu diperlukan. Sebagai akibatnya, kebutuhan informasi untuk sistem DW dapat dispesifikasi menggunakan model semantik untuk data multidimensional.

##### **b. Evaluasi skema data**

Skema data biasanya dikembangkan dengan metode khusus dan bukan oleh pengguna akhir, spesifikasi hasil kebutuhan harus dikaji oleh sasaran pengguna DW. Jika interpretasi kesenjangan signifikan atau

perubahan bisnis telah terjadi, maka harus kembali menganalisa permintaan informasi.

#### **3.14.5 Komponen Methodologi lainnya**

Terdapat beberapa metode komponen tambahan DW yang dapat dikembangkan yakni:

- a. Kumpulan jenis dokumen digunakan untuk menunjukkan hasil dari aktivitas tertentu.
- b. Kumpulan teknik yang digunakan untuk mendukung pembuatan dari dokumen dan / atau eksekusi dari aktivitas tertentu.
- c. Sebuah model organisasi yang menspesifikasikan aktivitas mana sajakah yang dilakukan dan diawasi oleh peran organisasi yang mana.
- d. Sebuah model informasi yang menspesifikasikan seluruh informasi obyek yang berhubungan untuk sebuah aktivitas, digambarkan dengan dokumen atau dibuat menggunakan teknik dan relasi antara informasi obyek tersebut.

#### **3.16 Requirement Specification**

Berikut ini adalah spesifikasi teknis yang diperlukan penulis untuk proses pengumpulan data kuis bahasa Inggris dan pembangunan DW:

1. Hardware:
  - a. Laptop Dell Inspiron5420.
  - b. Intel® Core™ i3-2370M CPU @ 2.40 GHz..
  - c. RAM 4.00 GB.
  - d. Windows 7 Home Basic
  - e. OS 64 Bit.

2. Software:

- a. ETL Tool: PDI Spoon.
- b. Cloud Storage: Google Drive.
- c. Questionare Tool: Google Form.
- d. Social Media: Facebook.
- e. Spreadsheet: Google Spreadsheet, Microsoft Excel.
- f. Grading Ad-On: Flubaroo.

3. Data:

- a. Input internal :
  - i. Data hasil responden kuis Bahasa Inggris.
  - ii. Biodata dari responden.
  - iii. Data waktu.
- b. Input external : Data wilayah diambil dari situs web Data.go.id
- c. Output : Fact hasil dengan format excel (.xlsx).

UMMN