

FASE-FASE YANG UMUM DIGUNAKAN DALAM PEMBANGUNAN DATA WAREHOUSE DAN KUALITAS DATA YANG HARUS DIPERTIMBANGKAN DI SETIAP FASENYA

Munawar

Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Esa Unggul, Jakarta
Jalan Arjuna Utara no.9, Tol Tomang, Kebon Jeruk, Jakarta Barat 11530
an_moenawar@yahoo.com

Abstrak

Many articles on data warehouse development have been written, however no clearly defined standards have been formulated that is applicable to all type of organizations. The rapid growth in data volumes has given rise to new problems for institutions: data quality, which is a critical issue when data are transferred from one system to another. Lack of data quality provided by data warehouse can lead to bad strategic decisions and indicates a significant failure rate. Thus, data quality in data warehouse needs to be assured. It has been widely accepted that data quality issues can emerge at any stage of data warehouse development. However, yet little work is done for formulating data quality that should be considered in the entirety of data warehouse development. This study was achieved through qualitative method by reviewing the most common practices data warehouse development in five organisations that is applicable to all type of organizations and then tried to confirm whether data quality in data warehouse from literature review is practiced in the five organisations followed by confirmation from the experts in order to determine specific data quality dimensions that correlated with data warehouse development. Based on the similarities in the development stages, identification of common practices for DW development can be obtained: requirements analysis, conceptual design, logical design, ETL, and physical design. There are sixteen dimensions of data quality that should be considered in the development of data warehouse.

Keywords : *data warehouse, data quality, common practices.*

Pendahuluan

Data Warehouse (DW) adalah gudang informasi (arsip) yang di kumpulkan dari berbagai sumber, yang disimpan di satu tempat di bawah skema terpadu (Silberschatz et.al, 2006). DW menyediakan antar muka kepada pengguna untuk konsolidasi data, tempat membuat penulisan query sehingga memudahkan user dalam melakukan pengambilan keputusan. Biasanya DW ini disimpan dalam bentuk dimensional (Rainardi, 2005).

Karena sifatnya yang sangat strategik, sangat penting untuk memastikan bahwa DW yang dibangun harus bisa dipastikan memenuhi standar kualitas dari semenjak fase awal pengembangannya (Rizzi, et. al., 2006). Semakin cepat masalah bisa teridentifikasi, semakin cepat rekomendasi bisa dibuat untuk memastikan bahwa data sudah dikoreksi dengan benar. Sayangnya, orang-orang yang terlibat

dalam pengembangan DW seringkali tidak mempertimbangkan dampak dari kualitas data (Kimball, et. al., 2008). Kualitas data yang baik akan memastikan tingkat kepercayaan pengguna DW sehingga bisa lebih diberdayakan yang akhirnya akan mengoptimalkan manfaat yang bisa didapatkan oleh dunia bisnis (Kumar dan Thareja, 2013). Dampaknya banyak proyek DW yang gagal karena kualitas data yang tidak baik (Hudicka, 2003). Hal ini disebabkan karena memastikan bahwa kualitas data yang digunakan di DW benar-benar bisa dijamin membutuhkan biaya yang besar serta waktu yang panjang (Nemoni and Konda, 2009). Karenanya kesuksesan pembangunan DW sangat tergantung kepada kepastian integrasi kualitas data ke semua fase pembangunan DW (Ballou & Tayi, 1999; Munawar, 2011). Jika tidak, kualitas data yang rendah akan mempengaruhi seluruh proses dan sistemnya

sendiri kalau sudah selesai di buat (Celko & McDonald, 1995).

Kualitas data di DW sangat penting, meski masih ada ketidaksepakatan tentang dimensi kualitas data apa yang perlu dimaksimalkan dalam pembangunan DW. Banyak memang organisasi yang sudah menyadari pentingnya kualitas data, namun perbaikan yang dilakukan masih sebatas ke akurasi data dan mengabaikan dimensi kualitas data yang lain (Prakash et al, 2004).

Selain perspektif obyektif, pandangan subyektif juga mempengaruhi dimensi kualitas data yang mana yang harus diadop, karena beda pengguna akan memformulasikan kriteria masing-masing sesuai dengan kebutuhan mereka. Karenanya, definisi terbaik tentang kualitas data adalah '*fit for use*' (Wang and Strong, 1996) dimana definisi ini menggarisbawahi sifat relatif dari data yang bisa jadi cocok pada suatu kondisi bisa jadi tidak cocok pada kondisi yang lain.

Motivasi Riset

Saat ini, semakin banyak organisasi yang membangun DW. Hal ini ditunjukkan dengan angka kenaikan pertahunnya 43 % (Winter and Strauch, 2003) dan akan selalu naik setiap tahunnya karena biaya *storage* yang semakin turun dari tahun ke tahunnya (Agosta, 2004). Namun sayang, dengan cepatnya pertumbuhan volume data telah menimbulkan problem baru bagi institusi yaitu kualitas data/ DQ (Ranjit and Kawaljeet, 2010), dimana hal ini adalah isu kritis ketika data ditransfer dari satu sistem ke sistem yang lain seperti sifat dasar DW.

Pengembangan dan pemeliharaan DW sangat mahal dan membutuhkan waktu panjang. Secara umum, pembangunan DW dalam skala penuh membutuhkan biaya lebih dari 1 Juta US Dollar (Hwang et al, 2004) dengan total biaya tahunannya berkisar 2-10 Juta US Dollar dimana hal ini termasuk dukungan staf untuk supervisi yang terdiri dari 20-50 staf (Amin and Arefin, 2010; Schiefer et al, 2002). Biaya rata-rata tahun pertama adalah 1.26 Juta US Dollar, dengan rata-rata waktu pembangunan 12-36 bulan (Amin and Arefin, 2010)

Hanya sayang, beberapa survey menunjukkan tingkat kegagalan pembangunan DW cukup tinggi yaitu 20-50% (Agosta, 2004; Conner, 2003; Watson et al, 2001) yang disebabkan oleh satu sebab yaitu ketiadaan kualitas (Cowie and Burstein, 2007). Data ini sekali lagi menunjukkan perlunya ada satu metode usaha yang didisain untuk menetapkan dimensi DQ yang kritis bagi pembangunan DW.

Rendahnya DQ akan menyebabkan biaya sebesar US\$600 milyar (Eckerson, 2002) dan mengkonsumsi kira-kira 10% dari keuntungan perusahaan (Redman, 2001). Rendahnya DQ juga akan meningkatkan biaya operasi dan efek ketidakpuasan pelanggan (Nemoni and Konda, 2009) dan menjadi kendala dalam penentuan keputusan bisnis yang tepat (English, 1999). Dampaknya akan menyulitkan eksekutif dalam menyelarkannya dengan tujuan dan obyektif organisasi. Lebih jauh, rendahnya DQ meningkatkan jumlah informasi dan pengulangan pekerjaan yang sia-sia yang berdampak kepada penghamburan tenaga, keuangan, material dan fasilitas (Pighin and Leronutti, 2008). Bahkan dalam banyak kasus akan menyebabkan ketidakefisienan (Redman, 2001) dan merupakan penyebab utama kegagalan (Haugh et al, 2011, Chenoweth et al, 2006; Hayen et al, 2007; Johnson, 2004; Ramamurthy et al, 2008).

Meskipun fenomena di atas menunjukkan pentingnya DQ di DW, namun hingga saat ini belum ada persetujuan bersama tentang DQ apa yang penting untuk DW. Sebuah konsensus diantara para peneliti menyatakan bahwa kualitas menyangkut beberapa dimensi (Ballou and Tayi, 1999; Cowie and Burstein, 2007; Jarke et al, 1999; Pipino et al, 2002; Prakash et al, 2004), namun tidak juga bisa menetapkan dimensi DQ apa saja yang penting yang perlu diakomodir oleh DW.

Kualitas adalah konsep kunci di pembangunan DW, dimana semua tahapnya berkaitan erat dengan pencapaian kualitas yang tinggi agar sukses diimplementasikan. Karenanya, pemahaman tentang dimensi DQ yang penting di setiap tahap pembangunan DW adalah hal yang kritis. Namun sepanjang yang peneliti ketahui, hampir tidak ada penelitian yang membahas dimensi DQ yang penting di

keseluruhan fase pembangunan DW. Pencarian dimensi DQ yang penting di semua fase pembangunan DW sangat penting untuk menjamin bahwa kualitas yang tinggi akan bisa dicapai (Nemoni and Konda, 2009).

Meski banyak peneliti yang menyetujui bahwa fase pembangunan DW mencakup hal-hal berikut: analisis kebutuhan, disain konseptual, disain logikal dan disain fisikal (Inmon, 2005; Kimball et al, 2008; Rizzi, 2009), namun penelitian empiris yang membahas masalah tersebut masih jarang. Beberapa klaim perlu dites secara empiris agar bisa diketahui fase-fase yang bisa dipakai di semua jenis organisasi. Oleh karena itu penting kiranya dilakukan penelitian tentang fase-fase yang bisa dipakai di semua jenis organisasi khususnya yang ada di Indonesia.

Saat ini relatif ada beberapa penelitian yang sudah dilakukan yang mengintegrasikan DQ ke fase pembangunan DW. Teknik DWQ digunakan untuk mengintegrasikan DQ ke fase desain dan penggunaan DW (Jarke et al, 1999), pertimbangan DQ di kebutuhan non fungsional dilakukan oleh Paim and Castro (2003), prosedur berbasis DQ untuk penggalian kebutuhan fungsional (queries) dan non fungsional dilakukan oleh Vaisman (2007), meta data berbasis model kualitas untuk memaksakan kualitas di DW dilakukan oleh Kumar and Thareja (2013) serta otomasi testing ETL terhadap DQ di DW (Dakrory et al, 2015) adalah beberapa contoh usaha yang sudah dilakukan untuk mengintegrasikan DQ di DW. Hanya sayang, meski semua teknik tersebut sudah memberikan pengetahuan tambahan terkait DQ di DW, namun kelemahan utamanya adalah analisis DQ pada teknik-teknik tersebut hanya dilakukan pada satu fase saja di pembangunan DW. Dengan demikian saat akan diintegrasikannya semua DQ ke semua fase pembangunan DW akan susah dilakukan. Sepanjang pengetahuan peneliti, hingga saat ini belum ada satupun penelitian yang mengusulkan pengintegrasian semua DQ yang penting di semua fase pembangunan DW. Hal inilah yang akan menjadi kontribusi penelitian ini.

Peningkatan DQ di DW sangat penting karena akan digunakan dalam mendukung proses pengambilan keputusan yang didukung oleh data yang akurat (Akbar et al., 2013).

Lemahnya DQ yang disiapkan oleh DW dapat mengakibatkan pengambilan keputusan strategik yang buruk. Oleh karena itu perlu usaha yang signifikan untuk memastikan bahwa DQ di DW benar-benar dijaga (Gosain and Heena, 2015). Informasi yang lebih baik, pengambilan keputusan yang lebih tepat hanya akan bisa didapatkan dari DQ yang tinggi di semua proses pembangunan DW (Verma et al, 2014). Semua itu hanya akan bisa dicapai melalui sebuah framework yang ditandai dengan diakomodirnya DQ ke dalam framework tersebut (Nemoni and Konda, 2009). Hanya sayangnya sepanjang pengetahuan peneliti, hingga saat ini belum ada satupun framework yang mengakomodir DQ di semua fase pembangunan DW. Oleh karena itu diharapkan dari penelitian ini akan bisa dijadikan sebagai bahan untuk penelitian berikutnya yaitu pembangunan framework DW yang mempertimbangkan DQ yang penting di semua fase pembangunan DW sehingga bisa meningkatkan kualitas DW yang dihasilkan.

Identifikasi Masalah dan Ruang Lingkup Penelitian

Isu kualitas data dapat terjadi dimanapun tahapan pada fase pembangunan DW (Kumar dan Thareja, 2013). Data sangat dipengaruhi oleh banyak proses yang harus dilalui hingga menuju ke DW. Keseluruhan proses ini akan mempengaruhi kualitasnya. Karenanya perlu diketahui dimensi kualitas apa yang akan menjamin kualitas data bisa didapatkan (Nemoni and Konda, 2009).

Meskipun isu kualitas data memiliki dampak langsung ke ekonomi dan sosial (Wang and Strong, 1996), namun masih sedikit penelitian yang dikembangkan untuk mengetahui kualitas data apa saja yang penting untuk diakomodir di semua fase pembangunan DW (Kumar, and Thareja, 2013).

Saat ini cukup banyak metodologi pengembangan DW yang ada di pasaran. Kelemahan utamanya adalah tidak bebas platform dan sangat tergantung pada vendor. Agar framework yang dihasilkan bisa dipakai di semua organisasi, penting untuk mengetahui praktek-praktek yang umum dipakai di banyak organisasi saat pembangunan data warehouse.

Dengan demikian akan bisa diterima oleh berbagai macam organisasi.

Tujuan Penelitian

Penelitian ini bertujuan untuk mengetahui fase-fase yang umum digunakan oleh organisasi saat membangun data warehouse agar fase – fase tersebut bisa diterima oleh semua jenis organisasi serta memahami kualitas data apa saja yang penting untuk diakomodir di semua fase pembangunan data warehouse.

Tahapan Penelitian

Tabel 1
Tahapan Penelitian

Tahapan	Tujuan
1. Studi literatur	1. Memahami fase-fase pembangunan data warehouse (DW) 2. Mengetahui aspek kualitas data (DQ) di data warehouse
2. Interview dan pengumpulan data	1. <i>Exploratory study</i> di organisasi yang sudah mengimplementasikan DW 2. <i>Expert judgement</i> oleh konsultan DW yang memiliki pengalaman dalam pembangunan DW minimal 3 siklus di sektor industri yang berbeda-beda
3. Analisis data	1. <i>Descriptive qualitative</i> digunakan untuk mendapatkan praktek umum yang biasa dilakukan oleh banyak organisasi dalam membangun DW sehingga bisa diadopsi oleh organisasi apapun. 2. Kuadran analysis digunakan untuk mengetahui dimensi kualitas Eppler yang penting diakomodir di semua fase pembangunan DW baik menurut studi empiris di banyak organisasi maupun oleh pakar

Pengumpulan Data

Mengingat besarnya biaya dan lamanya waktu yang dibutuhkan untuk pembangunan DW, bisa dipastikan hanya perusahaan besar sajalah yang mampu membangun DW. Dari sekian banyak perusahaan yang membangun DW di Indonesia, tidak semuanya bisa berpartisipasi dalam penelitian ini. Hal ini disamping karena kesibukan para pengambil keputusan (sehingga tidak memungkinkan untuk berpartisipasi) juga karena sifat DW yang memuat informasi yang sifatnya konfidensial sehingga tidak semua perusahaan bersedia untuk mengungkapkan hal-hal yang terkait dengan DW yang mereka bangun. Oleh karena itu, multi studi kasus (multi case study) adalah pendekatan yang paling tepat untuk pengumpulan data.

Manfaat Penelitian

Dengan adanya framework pembangunan DW yang mempertimbangkan DQ di semua fase pembangunan DW, bisa dijadikan pedoman bagi organisasi yang berniat membangun DW sehingga bisa mengurangi resiko kegagalan pembangunan DW.

Beberapa studi kasus akan bisa meminimalisir kelemahan satu studi kasus dalam mengeneralisir hasil temuan (Tellis, 1997). Beberapa studi kasus diharapkan akan bisa membantu mengetahui konsistensi dari beberapa kasus yang diteliti (Stake, 1995).

Setelah data dari beberapa studi kasus didapatkan, selanjutnya dilakukan konfirmasi hasil tersebut oleh beberapa pakar DW (konsultan DW yang berpengalaman lebih dari 7 tahun dengan kasus pembangunan DW yang sudah pernah dilakukan lebih dari 3 organisasi yang berbeda sektor industrinya).

Teknik penggalan informasi dilakukan dengan cara interview mendalam (*an in-depth interview*) untuk mendapatkan pemahaman yang teliti dan mendalam tentang *mengapa* dan *bagaimana* suatu kejadian bisa terjadi. Dari sini

selanjutnya hipotesis bisa dibuat dan ditest ke beberapa studi kasus (Flyvberg, 2006).

Interview dilakukan dengan menggunakan pertanyaan terstruktur untuk memastikan kelengkapan dan kemudahan direplikasikan di tempat lain. Tipe interview ini dipilih untuk mendapatkan data yang konsisten yang bisa dibandingkan ke semua responden. Pertanyaan terstruktur ini adalah cara pengumpulan data terbaik menurut literatur di area yang sedang di kembangkan sehingga memberikan pemahaman yang tepat kepada peneliti tentang suatu subyek tertentu. Selanjutnya akan memudahkan peneliti untuk membuat pertanyaan tertutup (*closed-ended questions*) yang relevan dan berarti bagi obyek penelitiannya (www.qualres.org/homeStru-3628.html). Untuk penelitian kualitatif, interview terstruktur ini sangat cocok untuk membandingkan respon yang diberikan oleh semua responden yang terdaftar terhadap pertanyaan yang sudah diberikan (Lindlof and Taylor, 2002).

Meski interview terstruktur ini memiliki banyak manfaat, namun memiliki kelemahan mendasar yaitu membutuhkan banyak sumberdaya (waktu dan uang). Meski demikian, teknik ini tetap dipilih karena kekhawatiran ketidakkonsistenan jawaban dari responden serta ketidakmengertian responden terhadap pertanyaan yang diberikan karena kompleksnya persoalan terkait dengan pembangunan DW.

Umumnya pertanyaan akan dimulai dengan pertanyaan terbuka kepada responden. Jawaban mereka akan dicatat dan selanjutnya diteruskan dengan kuesioner dengan menggunakan skor untuk mendapatkan jawaban kuantitatif sebagai manifestasi jawaban deskriptif yang sudah diberikan tentang situasi yang mereka hadapi. Skor ini digunakan untuk menyatakan tingkat setuju/ tidak setuju terhadap elemen penelitian dan juga tingkat kepuasan/ ketidakpuasan dari responden.

Interview dibagi menjadi dua group yang mewakili stakeholder yang berbeda yaitu : tim DW dan pengguna bisnis (*business users*). Tim DW mewakili stakeholder yang terkait dengan masalah teknik pembangunan DW, sedangkan pengguna bisnis mewakili pengguna yang menguasai proses bisnis dan tujuan organisasi yang terkait dengan DW. Pertanyaan

interview dibedakan antara tim DW dengan pengguna bisnis. Selanjutnya setiap responden diminta untuk mengisi kuesioner dengan menggunakan 7 skala untuk menunjukkan tingkat setuju/ tidak setuju terhadap elemen penelitian dan juga tingkat kepuasan/ ketidakpuasan dari responden. Coleho and Esteves (2007) menyatakan bahwa skala lebih dari 5 secara umum akan memberikan validitas hasil dan kemudahan penjelasan yang lebih baik dibanding yang menggunakan skala 5.

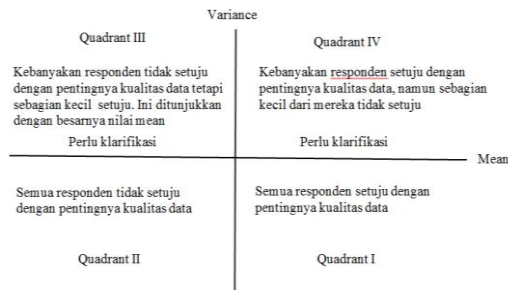
Selanjutnya data dari hasil kuesioner ini ditampilkan dalam bentuk tabel untuk mendukung temuan yang didapatkan dari hasil interview. Semua temuan dari interview dan kuesioner didiskusikan dan dianalisis secara mendalam dengan cara di review semua responden di organisasi yang berperan dalam penelitian ini dan juga semua konsultan yang terlibat untuk pengecekan kelengkapan dan keakuratannya.

Pengolahan Data

Karena sedikitnya jumlah ukuran sampel dan banyaknya variabel yang dipakai, maka digunakan gabungan antara penelitian deskriptif kualitatif dan kuantitatif sekaligus. Pendekatan deskriptif digunakan untuk identifikasi hubungan antar variabel yang terlibat sehingga bisa memberikan gambaran tentang fenomena yang ada, bahkan bisa digunakan untuk mengenali hubungan kausalitas diantara mereka (Punch, 2005).

Selanjutnya untuk memetakan respon responden di beberapa studi kasus di organisasi dan beberapa konsultan, dilakukan analisis kuadran. Jika diperlukan akan dilakukan juga sesi tambahan untuk mengklarifikasikan perbedaan perspektif yang ada di antara mereka.

Analisis kuadran ini sederhana untuk dilakukan dan dimengerti, namun sangat berguna untuk menganalisis kualitas informasi yang diperoleh, dimana terdapat dua kriteria untuk pengambilan keputusan. Kedua kriteria itu bisa dibedakan secara dikotomis (Richards dan Randolph, 2014).



Gambar 1

Analisis Kuadran untuk identifikasi data kualitas di setiap fase pembangunan DW

Hasil dan Pembahasan

Banyak metodologi pembangunan DW yang saat ini beredar di pasaran. Setiap metodologi ini dilengkapi dengan metodologi yang berbeda-beda satu dengan lainnya, yang lebih dikenal dengan metodologi khusus dari vendor. Kelemahan utama dari metodologi-metodologi ini adalah tidak netral dan tidak bebas platform. Oleh karenanya, dirasa sangat penting untuk mengetahui praktek-praktek yang umum dilakukan oleh organisasi saat membangun data warehouse agar bisa diterima oleh semua tipe organisasi.

Pada penelitian ini, praktek-praktek umum pembangunan DW didasarkan pada penelitian di 5 (lima) organisasi yang meliputi sektor industri yang berbeda yang terdiri dari : pemerintahan, asuransi, pendidikan, rumah sakit dan bank. Kelima organisasi ini juga mewakili tipe industri yang berbeda yaitu: pemerintahan

(1 organisasi), BUMN (2 organisasi) dan swasta (2 organisasi).

A. Fase yang Umum Dipakai dalam Pembangunan Data Warehouse

Dengan menggunakan instrumen interview yang berkaitan dengan pertanyaan apa saja input, proses dan output dari masing-masing fase pengembangan DW (sebagai contoh); semua organisasi yang diinterview menyatakan menggunakan goal-driven dan data-driven (n=5) saat penggalian kebutuhan untuk pembangunan DW. Mayoritas organisasi (n=3) menggunakan user-driven, dan sebagian kecil (n=2) menggunakan externally-driven dan process-driven saat penggalian kebutuhan. Berdasarkan hasil ini, lebih jauh dideskripsikan ke Tabel 4.1. Dengan cara yang sama digunakan untuk mengekstrak informasi terkait dengan fase-fase yang lain dalam pembangunan DW.

Analisis Kebutuhan

Semua organisasi yang diinterview menggunakan kombinasi dua atau lebih teknik analisis kebutuhan untuk menghindari kelemahan masing-masing pendekatan. Asuransi menggunakan lima pendekatan sekaligus yaitu *user-driven*, *goal-driven*, *externally-driven* dan *process-driven*. Sementara organisasi yang lain menggunakan 3 pendekatan. Deskripsi lebih rinci masing-masing organisasi bisa dilihat pada Tabel 2.

Tabel 2
Teknik Analisis Kebutuhan di Studi Kasus

Organisasi	User driven	Goal driven	Externally driven	Data driven	Process driven
Pemerintahan	√	√		√	
Bank		√	√	√	
Pendidikan	√	√		√	
Rumah Sakit		√		√	√
Asuransi	√	√	√	√	√

Disain Konseptual

Disain konseptual sangat penting dalam pembangunan DW karena memungkinkan pengembang DW mendeskripsikan kebutuhan dari sisi perspektif pengguna. Hal ini merepresentasikan komunikasi antara pengguna dan designer, meskipun tidak merefleksikan

secara mendalam fitur-fitur teknis untuk implementasinya.

Desain konseptual DW yang diadopsi oleh semua organisasi yang diinterview menggunakan teknik ER-relational dan atau ER-dimensional; tidak ada organisasi yang menggunakan teknik yang lain seperti pendekatan obyek oriented. Penentuan mana

yang terbaik bagi organisasi sangat susah ditentukan, namun yang sangat direkomendasikan adalah menggunakan ER-diagram meskipun skema logikalnya menggunakan relasional maupun dimensional. Berikut ini adalah deskripsi rinci hasil interview terkait dengan disain konseptual

Tabel 3
Pemodelan dimensional di Studi Kasus

Organisasi	M E/R	UML	Fact schema	DFM
Pemerintahan	√			
Bank	√			
Pendidikan	√			
Rumah Sakit	√			
Asuransi	√			

Disain Logikal

Skema star atau *snowflake* adalah representasi struktur logika dari DW. Skema ini menyiapkan view multidimensional yang memungkinkan untuk melakukan analisis data dalam perspektif yang berbeda. Analisis dilakukan dengan basis pengukuran (contoh jumlah produk terjual) dan dalam level yang berbeda untuk menunjukkan hirarki. OLAP membolehkan pengguna untuk melakukan agregasi dengan menggunakan fitur *roll-up* dan *drill-down*

Tabel 4
Disain Logikal di Studi Kasus

Organisasi	Implementasi Relational			Implementasi Multidimensional			Data storage		
	Star	Constellation	Snowflake	Condense	Dwarf	QC-Tree	ROLAP	HOLAP	MOLAP
Pemerintahan	√		√				√		
Bank	√						√		
Pendidikan	√		√				√		
Rumah Sakit	√						√		
Asuransi	√							√	

OLAP bisa disimpan dalam bentuk ROLAP, MOLAP dan HOLAP. Semua organisasi di studi kasus mendukung semua fungsi OLAP. Kompleksitas query dan kinerja adalah pertimbangan utama saat akan menggunakan ROLAP atau MOLAP. Gambaran lebih rinci dari hasil interview yang sudah dilakukan pada tahapan disain logikal bisa dilihat pada Tabel 4.

ekstrak dari sumber yang berbeda-beda, di transformasikan secara terpisah dan dimuat ke database DW. Sementara di ETL ekstraksi dan transformasi data dilakukan pada satu database. Lebih jauh implementasi ETL di studi kasus bisa dilihat pada Tabel 5.

ETL (Extract Transform Load)

Pertumbuhan volume data yang sangat dramatis pada era sekarang ini, menjadi tantangan tersendiri bagi organisasi yang mengimplementasikan DW. Pada fase ini, transformasi data menjadi tahapan yang paling menentukan. Keputusan apakah transformasi akan dilakukan di dalam atau di luar database DW sangat terkait dengan kinerja, kemudahan dan aspek keuangan. Di fase ETL, data di

Tabel 5
ETL di studi kasus

Organisasi	Ekstraksi			Transformasi			Loading		
	Transaction log	Immediate Database trigger	Source file	Deferred Time stamp	Manual File comparison	Manual Automated	Manual/Automated	Incremental	Full refresh
Pemerintahan				√		√		√	
Bank			√				√		√
Pendidikan				√			√	√	
Rumah Sakit				√			√	√	
Asuransi	√					√		√	

Disain Fisikal

Dua pendekatan yang umum digunakan dalam disain fisikal adalah top-down berdasarkan arsitektur DW tersentralisasi dan bottom-up dengan menggabungkan beberapa data mart ke sebuah data warehouse. Pada pendekatan pertama, data korporasi disiapkan oleh DW

yang tersentralisasi dan data mart yang bebas bisa didukung oleh DW tersentralisir untuk memenuhi kebutuhan unit yang membutuhkan. Sedangkan pada pendekatan kedua, DW dipresentasikan sebagai gabungan dari data mart yang berbeda-beda. Pendekatan lebih rinci pada studi kasus bisa dilihat pada Tabel 6.

Tabel 6
Arsitektur Data Warehouse di Studi Kasus

Organisasi	Data warehouse architecture				
	Enterprise DW	Data mart	Hub-and-spoke DM	Enterprise DW with ODS	Distributed DW
Pemerintahan		√			
Bank		√			
Pendidikan		√			
Rumah Sakit	√				
Asuransi		√			

Pertimbangan Kualitas Data dalam Pembangunan Data Warehouse

Semua organisasi yang diinterview menunjukkan ketertarikan dalam kaitan dengan kualitas data. Semua organisasi menyetujui bahwa kualitas data sangat penting dalam pembangunan DW. Kualitas data lebih menekankan kepada bagaimana dan kapan data digunakan untuk pengambilan keputusan. Kualitas data yang tinggi berkontribusi kepada organisasi dalam kemudahannya untuk dikonversikan ke pengetahuan melalui interpretasi dan evaluasi informasi dalam kaitannya dengan pengetahuan sebelumnya yang dibutuhkan dan juga melalui penggunaan informasi dalam konteks yang baru.

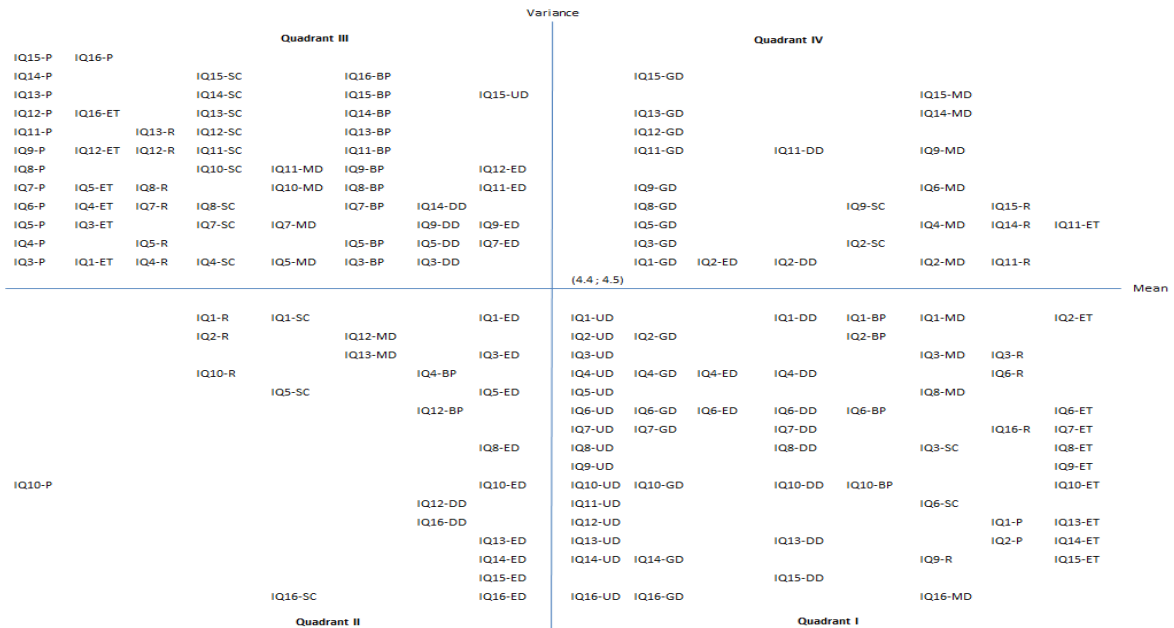
Untuk menjawab hal tersebut, respon kualitatif dari responden diubah menjadi kuantitatif melalui kuesioner secara tatap muka kepada responden. Secara keseluruhan, rata-rata mean dan variance adalah 4.4 dan 4.5. Nilai ini

digunakan sebagai titik pusat dari sumbu X (mean) dan sumbu Y (variance) untuk digunakan dalam analisis kuadran.

Kuadran analisis relatif sederhana untuk di eksekusi dan dipahami, meskipun sangat powerful dari sisi jumlah dan kualitas informasi yang dihasilkan. Dari hasil plotting ke dimensi kuadran bisa diketahui kualitas data yang penting, kurang penting dan seterusnya. Analisis kuadran ini menunjukkan posisi secara grafis 4 tipe kualitas data di dalam pembangunan DW. Kuadran I menunjukkan bahwa semua responden setuju atas pentingnya kualitas data di DW. Kuadran II menunjukkan bahwa semua responden sepakat tentang tidak pentingnya kualitas data di DW yang ditunjukkan oleh rendahnya nilai mean. Kuadran III menunjukkan bahwa mayoritas responden tidak setuju atas pentingnya kualitas data di DW. Klarifikasi dibutuhkan untuk menerangkan mengapa ada sedikit responden yang menyetujui

pentingnya kualitas data di DW pada kuadran ini. Kuadran IV merepresentasikan kebanyakan responden setuju atas pentingnya kualitas data di DW sebagaimana diindikasikan dengan tingginya nilai mean. Klarifikasi dibutuhkan untuk menjelaskan mengapa ada sedikit responden yang tidak menyepakati pentingnya

kualitas data di DW pada kuadran ini. Secara lebih rinci mean dan variance dari hasil kuesioner ini bisa dilihat pada Tabel 7, sedangkan kuadran kualitas data bisa dilihat pada Gambar 2.



Gambar 2
Kuadran Analisis Kualitas Data pada Studi Kasus

Tabel 7
Mean dan Variance dari Kualitas Data di Studi Kasus

		Data Warehouse Development Phase																					
		Requirements Analysis										Conceptual				Logical Design				ET		P	
		UD		GD		ED		DD		BP		MD		SC		R							
		Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var		
Information Quality	Content	Comprehensiveness	IQ1	5,2	1,7	4,8	5,2	3,5	3,7	5,2	1,7	4,6	5,3	4,5	3,0	4,0	4,0	3,2	5,2	4,0	3,0		
		Accuracy	IQ2	6,2	0,7	6,2	0,2	4,5	7,0	5,0	5,0	5,0	2,0	5,0	7,3	4,6	5,3	3,0	5,5	6,0	0,5	4,4	4,3
		Clarity	IQ3	6,0	0,5	5,4	6,3	4,0	2,7	3,0	7,5	3,2	6,7	5,0	4,0	5,2	3,2	5,2	3,2	3,4	6,3	3,0	5,5
		Applicability	IQ4	6,2	0,7	6,2	0,7	4,5	1,7	4,4	4,3	3,6	4,3	4,8	6,9	2,6	4,8	2,6	4,8	3,6	5,8	2,6	4,8
		Conciseness	IQ5	5,6	0,8	4,8	5,2	3,5	3,7	3,6	5,8	3,6	5,8	3,3	6,9	2,8	4,2	2,6	4,8	3,4	4,8	3,8	7,2
		Consistency	IQ6	6,2	0,2	6,2	0,2	4,5	3,0	6,3	0,3	6,4	0,3	5,0	7,3	4,8	2,7	4,6	2,3	5,6	0,3	3,8	6,7
	Process	Correctness	IQ7	5,6	0,3	4,6	4,3	3,8	4,9	6,0	1,0	4,0	8,0	3,3	6,9	3,0	7,5	4,0	7,5	5,8	0,2	2,8	6,2
		Currency	IQ8	6,0	1,0	5,2	6,2	4,3	2,3	6,4	0,3	3,8	6,7	5,5	1,0	3,6	6,3	3,4	5,3	5,2	0,7	3,4	5,3
		Convenience	IQ9	5,2	2,7	4,4	5,8	3,5	5,7	3,4	6,3	3,2	5,2	4,5	5,7	4,8	4,7	5,6	0,3	5,6	1,3	3,8	6,7
		Timeliness	IQ10	5,6	0,8	6,0	1,0	3,8	0,9	4,4	1,8	4,6	2,3	3,3	6,9	2,6	4,8	4,0	3,0	5,0	0,5	4,2	3,7
		Traceability	IQ11	5,6	1,3	5,2	6,2	4,0	6,7	6,0	5,0	4,0	9,0	4,0	12,0	3,2	9,2	5,4	6,3	5,6	6,8	4,2	8,7
		Interactivity	IQ12	4,6	2,8	4,4	4,8	3,0	5,3	3,0	4,0	2,8	4,2	4,0	4,0	3,4	4,8	3,6	5,8	2,6	4,8	2,6	4,8
	Infrastructure	Accessibility	IQ13	5,6	0,3	4,8	4,7	4,0	3,3	6,2	0,7	2,8	6,2	4,3	4,9	4,0	8,0	4,2	8,7	5,4	0,3	2,8	6,2
		Security Information	IQ14	6,8	0,2	6,6	0,3	3,5	4,3	4,2	8,7	3,0	7,5	4,8	6,3	3,8	6,7	5,0	5,5	4,4	4,3	3,8	6,7
		Information Maintainability	IQ15	4,2	6,2	5,0	5,0	3,8	2,9	6,4	0,3	3,6	6,3	4,5	5,7	4,2	4,7	4,4	5,3	5,6	0,8	4,2	4,7
		Speed	IQ16	5,2	2,2	5,4	0,8	2,8	4,3	4,2	2,7	3,8	6,7	5,3	1,6	4,2	3,7	4,4	2,3	3,4	5,3	3,8	6,7

Berdasarkan analisis kuadran seperti pada Gambar 2, pentingnya kualitas data di DW bisa digambarkan lebih lanjut. Tabel 8 menunjukkan pentingnya kualitas data di semua fase pembangunan DW. Ada beberapa perbedaan persepsi diantara responden (lihat tanda ♦♦). Beberapa perbedaan bisa dilihat

(sebagai contoh currency (masih berlaku) – IQ8). Currency kurang penting di pemerintahan; namun sangat penting buat BUMN dan organisasi swasta.

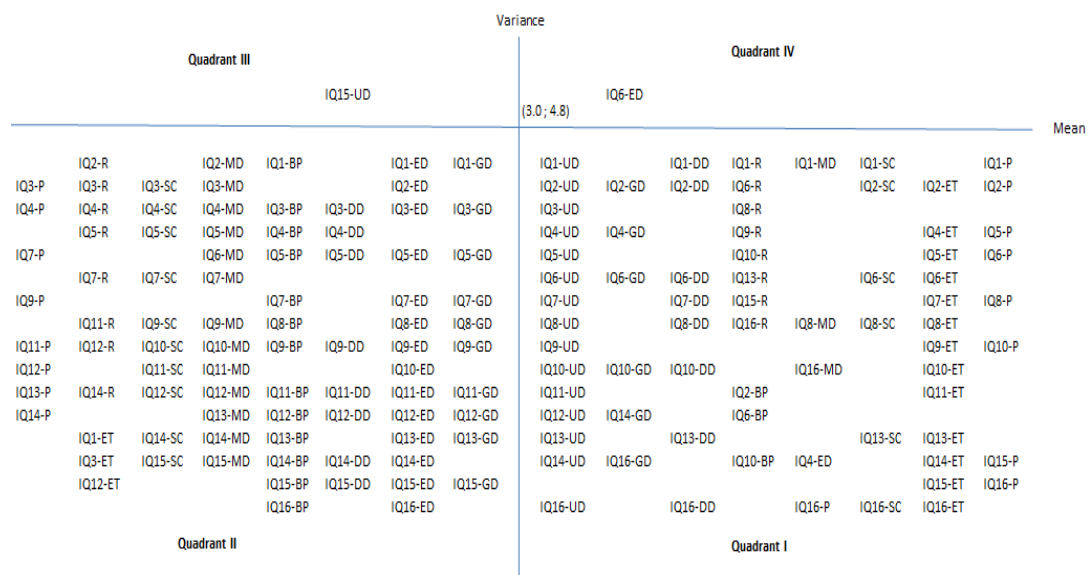
Pertanyaan yang sama selanjutnya ditujukan kepada pakar untuk memvalidasi kualitas data yang ditemukan pada studi kasus.

Tiga konsultan DW diminta menganalisis hasil tersebut dan diminta untuk mengkonfirmasi temuan pada studi kasus yang sudah dilakukan. Hasil lebih rincinya bisa dilihat pada Tabel 9 dan Gambar 3.

Tabel 8
Pentingnya Kualitas Data di Studi Kasus

◆ All organizations agree on the importance of DQ
◆◆ More than 3 organizations agree on the importance of DQ

			Variables of Data Warehouse Development												
			Requirements					Conceptual	Logical		ETL	Physical			
			User Driven	Goal Driven	External Driven	Data-Driven	Process-driven	Multi Dimensional	Star Schema	OLAP	ETL	Physical Design			
			UD	GD	ED	DD	BP	MD	SC	R	ET	P			
Data Quality Dimensions Variable	Information Quality	Content	Relevance	Comprehensiveness	IQ1	◆	◆◆		◆	◆	◆				◆
			Accuracy	IQ2	◆	◆	◆◆	◆◆	◆	◆◆	◆◆		◆	◆	
			Clarity	IQ3	◆	◆◆				◆	◆	◆			
			Applicability	IQ4	◆	◆	◆	◆		◆◆					
		Soundness	Conciseness	IQ5	◆	◆◆									
			Consistency	IQ6	◆	◆	◆	◆	◆	◆◆	◆	◆	◆		
			Correctness	IQ7	◆	◆		◆						◆	
			Currency	IQ8	◆	◆◆		◆		◆				◆	
	Access	Process	Convenience	IQ9	◆	◆◆				◆◆	◆◆	◆	◆		
			Timeliness	IQ10	◆	◆		◆	◆					◆	
			Traceability	IQ11	◆	◆◆		◆◆				◆◆	◆◆		
			Interactivity	IQ12	◆	◆◆									
		Infrastructure	Accessibility	IQ13	◆	◆◆		◆						◆	
			Security Information	IQ14	◆	◆				◆◆		◆◆	◆	◆	
			Information Maintainability	IQ15		◆◆		◆		◆◆		◆◆	◆		
			Speed	IQ16	◆	◆				◆		◆			



Gambar 3
Kuadran Analisis Kualitas Data pada Expert Judgement

Berdasarkan gambar 3 terlihat pandangan pakar. Para konsultan secara umum pentingnya kualitas data di DW menurut menyetujui pentingnya penentuan kualitas data

terkait dengan proses bisnis dan pengambilan keputusan. Dalam kaitan dengan hal-hal teknis (konseptual, logikal, ETL dan fisik) terjadi perbedaan perspektif antar pakar dan studi kasus. Meskipun pada studi kasus, orang-orang IT sudah membantu dalam menentukan kualitas data yang dibutuhkan di pembangunan DW, namun validasi dari para pakar menunjukkan beberapa perbedaan. Sebagai contoh accuracy (IQ2) menurut para pakar tidak relevan dalam kaitannya dengan pihak luar sementara menurut

organisasi yang disurvei, accuracy penting dalam kaitannya dengan pihak luar. Faktanya, kadangkala informasi yang dibutuhkan oleh pihak luar tidak tersedia di sumber data di organisasi. Oleh karenanya lebih tepat jika dikatakan consistency (IQ2) yang relevan bagi pihak luar dibanding dengan consistency. Lebih jauh perbandingan antara pakar dengan pengguna di studi kasus bisa dilihat pada Tabel 10.

Tabel 9
Mean dan Variance dari Kualitas Data dari Expert Judgement

		Data Warehouse Development Phase																					
		Requirements Analysis										Conceptual				Logical Design				ET		P	
		UD		GD		ED		DD		BP		MD		SC		R		ETL		Physical Design			
		User Driven	Goal Driven	Externally Driven	Data Driven	Process Driven	Multi Dimensional	Star Schema	OLAP	ETL	Physical Design												
		Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var		
Information Quality	Content	Comprehensiveness	IQ1	5,3	1,3	1,0	0,0	1,0	0,0	5,0	0,0	1,0	0,0	5,0	0,0	5,0	0,0	1,0	0,0	5,0	0,0		
		Accuracy	IQ2	6,7	0,3	6,0	0,0	1,0	0,0	6,0	0,0	4,7	0,3	1,0	0,0	5,3	2,3	1,0	0,0	6,7	0,3	4,7	0,3
		Clarity	IQ3	6,0	1,0	1,0	0,0	2,0	0,0	1,0	0,0	1,0	0,0	2,0	0,0	2,0	0,0	2,0	0,0	1,0	0,0	2,0	0,0
		Applicability	IQ4	5,7	1,3	5,3	0,3	3,3	4,3	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	6,7	0,3	1,0	0,0
		Conciseness	IQ5	5,7	1,3	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	5,0	0,0	7,0	0,0
		Consistency	IQ6	6,7	0,3	6,0	0,0	3,7	5,3	7,0	0,0	7,0	0,0	1,0	0,0	3,3	1,3	3,3	1,3	6,7	0,3	6,0	0,0
		Correctness	IQ7	6,3	0,3	1,0	0,0	1,0	0,0	7,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	6,0	0,0	1,0	0,0
		Currency	IQ8	6,0	1,0	1,0	0,0	1,7	1,3	7,0	0,0	1,0	0,0	4,0	0,0	4,0	0,0	4,0	0,0	4,0	0,0	4,0	0,0
	Access	Convenience	IQ9	4,7	1,3	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	5,0	0,0	4,7	0,3	1,0	0,0		
		Timeliness	IQ10	5,3	0,3	6,0	1,0	2,3	1,3	3,3	0,3	5,7	0,3	1,0	0,0	1,0	0,0	4,3	0,3	6,0	0,0	4,3	0,3
		Traceability	IQ11	4,7	4,3	1,7	1,3	1,0	0,0	2,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	5,3	0,3	1,0	0,0
		Interactivity	IQ12	3,7	1,3	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0	1,0	0,0
		Accessibility	IQ13	6,0	1,0	1,0	0,0	2,3	1,3	7,0	0,0	1,0	0,0	1,0	0,0	7,0	0,0	7,0	0,0	5,0	0,0	1,0	0,0
		Security Information	IQ14	7,0	0,0	7,0	0,0	1,7	1,3	2,3	5,3	1,0	0,0	1,3	0,3	1,3	0,3	1,3	0,3	4,0	0,0	1,0	0,0
		Information Maintainability	IQ15	3,3	3,6	1,0	0,0	2,3	1,3	1,0	0,0	1,0	0,0	1,0	0,0	3,3	0,3	3,3	0,3	6,7	0,3	3,3	0,3
		Speed	IQ16	6,0	0,0	4,7	0,3	1,0	0,0	3,0	0,0	1,0	0,0	4,0	0,0	4,0	0,0	4,0	0,0	5,7	0,3	5,0	0,0

Tabel 10
Perbandingan antara Kualitas Data di Studi Kasus dan Pakar

- ◆ All respondents agree on the importance of DQ
- ◆◆ Most of respondents agree on the importance of DQ
- CS : Case Studies
- EX : Expert Judgment

		Variables of Data Warehouse Development																							
		Requirements										Conceptual				Logical				ETL		Physical			
		User Driven		Goal Driven		External Driven		Data-Driven		Process-driven		Multi Dimensional		Star Schema		OLAP		ETL		Physical Design					
		CS	EX	CS	EX	CS	EX	CS	EX	CS	EX	CS	EX	CS	EX	CS	EX	CS	EX	CS	EX				
Data Quality Dimensions Variable	Information Quality	Content	Comprehensiveness	IQ1	◆	◆	◆◆			◆	◆	◆		◆	◆		◆				◆	◆			
			Accuracy	IQ2	◆	◆	◆	◆◆			◆◆	◆	◆	◆	◆◆		◆◆	◆			◆	◆	◆	◆	
			Clarity	IQ3	◆	◆	◆◆								◆		◆								
			Applicability	IQ4	◆	◆	◆	◆	◆	◆	◆	◆			◆◆								◆		
		Soundness	Conciseness	IQ5	◆	◆	◆◆															◆		◆	
			Consistency	IQ6	◆	◆	◆	◆	◆	◆◆	◆	◆	◆	◆	◆	◆◆		◆	◆	◆	◆	◆	◆	◆	◆
			Correctness	IQ7	◆	◆	◆				◆	◆										◆	◆		
			Currency	IQ8	◆	◆	◆◆								◆	◆		◆				◆	◆	◆	◆
	Access	Convenience	IQ9	◆	◆	◆◆							◆◆		◆◆		◆	◆	◆	◆	◆	◆	◆	◆	
		Timeliness	IQ10	◆	◆	◆	◆														◆	◆	◆	◆	
		Traceability	IQ11	◆	◆	◆◆				◆◆							◆◆		◆◆	◆					
		Interactivity	IQ12	◆	◆	◆◆																			
		Accessibility	IQ13	◆	◆	◆◆				◆	◆					◆		◆	◆	◆	◆				
		Security Information	IQ14	◆	◆	◆	◆							◆◆		◆◆		◆	◆	◆	◆				
		Information Maintainability	IQ15			◆◆				◆				◆◆		◆◆		◆	◆	◆	◆			◆	
		Speed	IQ16	◆	◆	◆	◆							◆	◆		◆	◆	◆	◆	◆			◆	

Berdasarkan hasil interview di 5 organisasi dan diskusi dengan 3 pakar bisa disimpulkan bahwa tahapan-tahapan yang umum dilakukan oleh organisasi dalam membangun DW meliputi hal-hal sebagai berikut: analisis kebutuhan (goal-driven, data-driven, externally-driven, process-driven dan user-driven), disain konseptual, disain logical (star schema dan OLAP), ETL dan disain fisik. Tahapan-tahapan ini akan bisa dipakai oleh semua organisasi tanpa membedakan teknologi dan metodologi yang digunakan.

Dalam kaitannya dengan penentuan kualitas data yang penting untuk diakomodir di semua fase pembangunan DW ada perbedaan diantara para pengguna di studi kasus dan para pakar. Secara umum bisa dikatakan bahwa terkait dengan masalah teknis (konseptual, logikal, ETL dan fisik) pandangan pakar lebih mendekati kondisi nyata yang ada di lapangan. Sementara dalam kaitan dengan proses bisnis, pandangan dari para pengguna dalam kasus tertentu akan lebih sesuai dengan kondisi faktual. Dengan demikian justifikasi dari pengguna sangat diperlukan dalam kasus-kasus yang khusus.

Kesimpulan

Data warehouse adalah sistem yang kompleks yang membutuhkan kerja sama antar pelaku bisnis dan pengembang data warehouse. Berdasarkan literatur terlihat bahwa kegagalan proyek DW banyak diakibatkan karena kualitas datanya kurang terpenuhi. Atas dasar alasan inilah penelitian ini dilakukan.

Langkah pertama yang perlu dilakukan untuk menjawab persoalan yang sudah dikemukakan di atas adalah dengan mencari praktek-praktek umum yang biasa dilakukan oleh banyak organisasi saat membangun data warehouse. Dengan teknik studi eksplorasi di lima organisasi bisa didapatkan tahapan-tahapan yang umum dipakai di organisasi sehingga bisa dipakai di berbagai tipe organisasi. Tahapan yang umum tersebut terdiri atas : analisis kebutuhan (goal-driven, data-driven, externally-driven, process-driven dan user- driven), disain konseptual, disain logical (star schema dan OLAP), ETL dan disain fisik.

Tahap berikutnya yang perlu dilakukan adalah menentukan kualitas data apa saja yang harus diakomodir saat membangun data warehouse. Studi di lima organisasi dan juga konfirmasi dari 3 pakar pembangunan data warehouse menunjukkan bahwa kualitas data bersifat subyektif. Hal ini tergantung pada kepentingan pengguna dan tipe industri nya. Semua kualitas data yang ada di literatur tidak semuanya sesuai dengan kondisi aktual di lapangan. Perlu kolaborasi antara pengguna, orang IT dan juga pakar pengembang data warehouse untuk menentukan kualitas data apa saja yang memang diperlukan saat membangun data warehouse.

Karena keterbatasan waktu para eksekutif untuk diwawancarai, maka pada penelitian ini hanya bisa dilakukan studi kasus di lima organisasi. Karena itu penelitian ini akan bisa memberikan hasil yang lebih akurat jika lebih banyak organisasi yang bisa terlibat.

Agar lebih berdaya guna, penelitian berikutnya perlu dilakukan untuk melihat efektifitas pengintegrasian kualitas data ke semua fase pembangunan data warehouse terhadap benefit yang diberikan. Dengan demikian biaya yang lumayan besar untuk memastikan kualitas data yang tinggi bisa sepadan dengan manfaat yang akan diperoleh.

Daftar Pustaka

- Agosta, L. (2004). Data Warehousing Lessons Learned: A Time of Growth for Data Warehousing, in *DM Review Magazine*, 2004, pp. Retrieved on 29/3/2011, from http://www.dmreview.com/article_sub.cf?articleId=1012461.
- Akbar, K., Krishna, S.M and Reddy, V.R. (2013). ETL Process Modeling In DWH Using Enhanced Quality Techniques. *International Journal of Database Theory and Application* Vol. 6, No. 4, August, 2013.
- Al-Debei, M.M. (2011). Data Warehouse as a Backbone for Business Intelligence. *European Journal of Economics, Finance, and Administrative Sciences*, 33, 153-166.

- Alshawi, S., Saez-Pujol, I. and Irani, Z. (2003). Data warehousing in decision support for pharmaceutical R & D supply chain. *International Journal of Information Management*, 23, 259-268. [http://dx.doi.org/10.1016/S0268-4012\(03\)00028-8](http://dx.doi.org/10.1016/S0268-4012(03)00028-8)
- Amin, M.R and Arefin, M.T. (2010). The Empirical Study on the Factors Affecting Data Warehousing Success. *International Journal of Latest Trends in Computing (E-ISSN: 2045-5364) Volume 1, Issue 2, December 2010*
- Babbie, E. R. (2009). *The Practice of Social Research*, 12th edition, Wadsworth Publishing, 2009, ISBN 0-495-59841-0, pp. 436–440
- Ballou, D. P., and Tayi, G.K. 1999. Enhancing Data Quality in Data Warehouse Environments. *Communications of the ACM* January 1999/Vol. 42, No. 1. pp 73-78
- Bilal Ali, Y. A. (2014). Challenges in the Successful Implementation of Data Warehouse. *Journal of Management Research*. ISSN 1941-899X. 2014. Vol 6, No 3
- Celko, J., & McDonald, J. 1995. Don't warehouse dirty data. *Datamation*, 41(19), 42–53.
- Chenoweth, T., Corral, K. and Demirkan, H. (2006). Seven key interventions for data warehouse success, *Communications of the ACM*, vol. 49, pp. 114-119.
- Conner, D. (2003). Data warehouse failures commonplace, *Network World*, vol. 20, p. 24.
- Cowie, J. and Burstein, F. (2007). Quality of data model for supporting mobile decision making. *Decision Support Systems* 43, 1675–1683
- Coleho, P.S. and Esteves, S.P. (2007). The choice between a 5-point and 10-point scale in the framework of customer satisfaction measurement. *International Journal of Market Research*, Vol. 49. No. 3, 2007, pp.313-345
- Dakrory, S.B, Mahmoud, T.M., and Ali, A.A. (2015). Automated ETL Testing on the Data Quality of a Data Warehouse. *International Journal of Computer Applications* 131(16):9-16, December 2015. Published by Foundation of Computer Science (FCS), NY, USA
- Eckerson, W (2002) *Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data*. The Data Warehousing Institute, Seattle, WA.
- Eppler, M. J. (2006). *Managing Information Quality: Increasing the Value of Information in Knowledge-Intensive Products and Processes* (2nd ed.): Springer.
- Flyvberg, B. (2006). Five misunderstandings about case study research. *Qualitative Inquiry* 12, 2, 219-245
- Gosain, A. and Heena. (2015). Literature Review of Data model Quality metrics of Data Warehouse. *International Conference on Computer, Communication and Convergence (ICCC 2015)*. doi: 10.1016/j.procs.2015.04.176
- Griffin, R. K. (1998). Data Warehousing. *Cornell Hospitality Quarterly*, 39, 28-35.
- Hwang, H. G., Ku, C. Y., Yen, D. C. and Cheng, C. C. (2004). Critical factors influencing the adoption of data warehouse technology: a study of the banking industry in Taiwan, *Decision Support Systems*, vol. 37, pp. 1-21.
- Hwang, M. I., & Xu, H. (2008). A Structural Model of Data Warehousing Success.

Journal of Computer Information Systems, 49(1).

- Hudicka, J. 2003. Bumpy Ride - Data Migration Projects Still Plagued by Problems. *Intelligent Enterprise*, 10
- Haug, A., Zachariassen, F and Van Liempd, D. (2011). The Costs of Poor Data Quality. *Journal of Industrial Engineering and Management (JIEM 2011)*, 4(2): 168-193
- Hwang, H. G., Ku, C. Y., Yen, D. C. and Cheng, C. C. (2004). Critical factors influencing the adoption of data warehouse technology: a study of the banking industry in Taiwan, *Decision Support Systems*, vol. 37, pp. 1-21.
- Hayen, R. L., Rutashobya, C. D. and Vetter, D. E. (2007). An Investigation of the Factors Affecting Data Warehousing Success, *International Association for Computer Information Systems (IACIS)*, vol. 8, pp. 547-553.
- Inmon, W. H. (2005). Building the data warehouse (4th ed.): John Wiley & Sons, Inc. New York, NY, USA.
- Jaklic, J., Coelho, P. S. And Popovic, A. (2009). Information Quality Improvement as a Measure of Business Intelligence System Benefits. *WSEAS Transactions on Business and Economics*. Issue 9. Volume 6. September 2009. ISSN. 1109-9526
- Jarke, M., Jeusfeld, M., Quix, C., Vassiliadis, P. (1999). Architecture and Quality in Data Warehouses: An Extended Repository Approach. *Information Systems* 24(3), 229–253
- Johnson, L. K. (2004). Strategies for Data Warehousing, *MIT Sloan Management Review*, vol. 45, p. 9.
- Joseph, M.V. (2013). Significance of Data Warehousing and Data Mining in Business Applications. *International Journal of Soft Computing and Engineering (IJSCE)* ISSN: 2231-2307, Volume-3, Issue-1, March 2013
- Kimball, R., Ross, M., Thornthwaite, W., Mundy, J. and Becker, B. (2008). *The Data Warehouse Life Cycle Toolkit: Practical Techniques for Building Data Warehouse and Business Intelligent Systems*. Second Edition. Wiley Publishing, Inc. IN 46256
- Kumar, V. and Thareja, R. (2013). A Simplified Approach for Quality Management in Data Warehouse. *International Journal of Data Mining & Knowledge Management Process (IJDMP)* Vol.3, No.5, September 2013
- Lindlof and Taylor. (2002). *Qualitative Communication Research Methods*, 2nd Edition. Thousands Oaks: SAGE. ISBN 978-0-7619-2494-4
- Loshin, D. (2008). The Data Quality Business Case: Projecting Return on Investment, Informatica White paper. Retrieved July 15, 2010 from <http://www.melissadata.com/enews/articles/1007/2.htm>
- Mukherjee, D. (2003). An empirical investigation of critical factors that influence data warehouse implementation success.
- MajidZaman, E., Quadri, S.M.K., and But, M. A. (2012). Information Integration for Heterogeneous Data Sources. *IOSR Journal of Engineering* Apr. 2012, Vol. 2(4) pp: 640-643. ISSN: 2250-3021
- Munawar, Salim, N., Ibrahim, R. (2011). Toward Data Quality Integration into the Data Warehouse Development. In: Ninth IEEE International Conference on Dependable, Autonomic and Secure Computing. IEEE Computer Society, 978-0-7695-4612-4/11, doi:10.1109/DASC.2011.194

- Nemoni, R. R and Konda, R. 2009. A Framework for Data Quality in Datawarehouse. In J. Yang et. Al (Eds): UNISCON 2009, LNBIP 20, pp 292 – 297. Springer-Verlag Berlin Heidelberg 2009
- Paim, F. R. S., and Castro, J. B. (2003). DWARF: An Approach for Requirements Definition and Management of Datawarehouse Systems. In Proceedings of the 11th IEEE International Conference on Requirement Engineering (pp 75-78)
- Pandey, A. and Mishra, S. (2014). Moving from Traditional Data Warehouse to Enterprise Data Management : a Case Study. Issues in Information Systems Volume 15, Issue II, pp. 133-140, 2014
- Prakash, N., Singh, Y., Gosain, A. (2004). Informational Scenarios for Data Warehouse Requirements Elicitation. In: Atzeni, P., Chu, W., Lu, H., Zhou, S., Ling, T.-W. (eds.) ER 2004. LNCS, vol. 3288, pp. 205–216. Springer, Heidelberg
- Pighin and Leronutti, (2008). A Methodology Supporting the Design and Evaluating the Final Quality of Data Warehouses. IGI Global
- Pipino, L., Lee, Y., Wang, R. (2002). Data Quality Assessment. Commun. ACM 45, 4
- Punch, K. F. (2005). *Introduction to Social Research: quantitative and qualitative approaches*, London: SAGE.
- Ramamurthy, K., Sen, A., and Sinha, A. P. (2008). An empirical investigation of the key determinants of data warehouse adoption, *Decision Support Systems*, vol. 44, pp. 817-841.
- Que, W. T. (1988). *Marketing Research. Marketing Research of Singapore*, 1988. ISBN 981-00-0610-1, pp. 385-388
- Rahman, N., Marz, J., and Akhter, S. (2012). An ETL Metadata Model for Data Warehousing. *Journal of Computing and Information Technology - CIT* 20, 2012, 2, 95–111. doi:10.2498/cit.1002046
- Rainardi, Vincent. (2008). *Building a data warehouse: with examples in sql server*. A Press.
- Ranjit Singh and Kawaljeet Singh. (2010). A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing. *IJCSI International Journal of Computer Science Issues*. Vol. 7, Issue 3, No 2 May 2010. ISSN : 1694-0784
- Redman, T.C. (2001). *Data Quality: The field guide*. Digital Press, Boston.
- Richards J. Heur Jr and Randolph H. Pherson (2014). *Structured Analytic Techniques for Intelligent Analysis*. CQ Press.
- Rizzi, S., Abelló, A., Lechtenböcker, J., & Trujillo, J. 2006. Research in data warehouse modeling and design: Dead or alive? In *Proceedings of the 9th ACM Int. Workshop on Data Warehousing and OLAP (DOLAP '06)*, (pp. 3-10) ACM Press
- Rizzi, S. (2009). *Conceptual Modelling Solutions for the Data Warehouse*. IGI Global
- Schiefer, J., List, B. and Bruckner, R.M. (2002). A Holistic Approach for Managing Requirements of Data Warehouse Systems. Eight Americas Conference on Information Systems.
- Shahzad, M.K. (2012). *Improving Business Processes using Processoriented Data Warehouse*. Doctoral Dissertation in Computer and Systems Sciences School of Information and Communication Technologies KTH - Royal Institute of Technology, Stockholm, Sweden

- Silberschatz, Avi., Hank Korth, S. Sudarshan. (2006). *Database System Concepts, Fifth Edition*. New York : McGraw-Hill.
- Stake, R. E. (1995). *The Art of Case Study Research*, London: SAGE.
- Summer, E., & Ali, D. (1996). A Practical Guide for Implementing Data Warehousing. *Computers Ind. Engng*, 31, 307-310.
[http://dx.doi.org/10.1016/0360-8352\(96\)00137-4](http://dx.doi.org/10.1016/0360-8352(96)00137-4)
- Tellis, W. (1997). Introduction to case study. *The Qualitative Report*, 3 (2), 1-11.
- Vaisman, A. (2007). Data Quality-Based Requirements Elicitation for Decision Support Systems. in *Data Warehouse and OLAP : Concepts, Architectures and Solutions*. pp 58-86. M. Gordon Hunter (University of Lethbridge, Canada). DOI: 10.4018/978-1-60566-090-5
- Verma, D, Tyagi, A., and Sharma D. (2014). Data Quality Problems in Data Warehousing. *International Journal of Innovative Research in Technology*. Volume 1 Issue 7. ISSN: 2349-6002
- Wang, R.Y. and Strong, D.M. (1996). Beyond accuracy: What data quality means to data consumers, *Journal of Management Information Systems*. 12(4): 5-34.
- Watson, H.J. and Haley, B.J. (1997). Data warehousing: a framework and survey of current practices, *Journal of Data Warehousing* 2 (1) (1997) 10-17.
- Watson, H. J., Goodhue, D. L. and Wixom, B. H. (2002). The benefits of data warehousing: why some organizations realize exceptional payoffs. *Information and Management*, vol. 39, pp. 491-502.
- Watson, H. and Haley, B. (1998). A Structural Model of Data Warehousing Success. *Journal of Data Warehousing*, 2, 10-17.
- Winter, R and Strauch, B. (2003). A method for demand-driven information requirements analysis in data warehousing. In *Proceedings of Hawaii International Conference on System Sciences*, Hawaii, 1359-1365
- Wixom, B. H. and Watson, H. J. (2001). An Empirical Investigation of the Factors Affecting Data Warehousing Success. *MIS Quarterly*, vol. 25, pp. 17-41