

PERANCANGAN DATA WAREHOUSE PENJUALAN (STUDI KASUS PT. SUBAFOOD PANGAN JAYA)

Dian Sugiarto¹, Harco Leslie Hendric Spits Warnars², Winarno³

*Magister Teknik Informatika, Perguruan Tinggi Raharja
Jl. Jenderal Sudirman No.40, Modern Cikokol, Tangerang
dian.sugiarto@raharja.info*

*Computer Science Department, BINUS Graduate Program –Doctor of Computer Science
Bina Nusantara University, Jakarta, Indonesia 11480
spits.hendric@binus.ac.id*

*Direktur LPPM UMN
Universitas Multimedia Nusantara, Tangerang, Indonesia
pmwinarno@unimedia.ac.id*

ABSTRAK

Penjualan merupakan suatu hal yang penting bagi keberlangsungan hidup suatu perusahaan. Kegiatan penjualan merupakan kegiatan utama perlu untuk dikelola secara baik agar perusahaan tidak selalu mengalami kerugian. Pada penelitian ini penulis akan menoba untuk melakukan perancangan *Data Warehouse* yaitu sebuah lokasi penyimpanan data yang mempunyai ukuran yang sangat besar, bersifat historis dan dapat dipergunakan untuk memberikan dukungan kepada pimpinan perusahaan dalam mengambil keputusan. Dalam penelitian ini juga akan merancang sebuah *Data Warehouse* sebagai *repository* penjualan yang diterapkan menggunakan *Pentaho Data Integration Software*. Hasil dari Penelitian ini adalah Sebuah rancangan *Data Warehouse* yang dapat digunakan sebagai *repository* data-data penjualan.

Kata Kunci: Perancangan, *Data Warehouse*, *Pentaho Data Integration*, *Repository*.

ABSTRACT

Sales is an important thing for the survival of a company. Sales activities are the main activities that must be managed properly so that the company does not always suffer losses. This study will design the Data Warehouse, which is a data storage location that has a very large size, is historical and can be used to support company leaders in making decisions. In this study also will design a Data Warehouse as a sales repository that is implemented using Pentaho Data Integration Software. The results of this study are a Data Warehouse design that can be used as a repository of sales data.

Keywords: Design, Data Warehouse, Pentaho Data Integration, Repository.

PENDAHULUAN

Pesatnya perkembangan teknologi informasi saat ini telah membuat dunia pemerintahan, akademik, bisnis dan berbagai macam bidang lainnya untuk memanfaatkan perkembangan teknologi tersebut secara maksimal. Banyak proses bisnis dan pengolahan data sangat bergantung dengan keberadaan teknologi informasi. Hal ini telah mendorong perusahaan-perusahaan yang ingin berkompetisi atau memenangkan persaingan melakukan terobosan-terobosan terkait pemanfaatan teknologi informasi itu sendiri.

Informasi dan data yang dikelola atau dimiliki oleh perusahaan kini telah berubah menjadi aset

berharga untuk kepentingan kelangsungan dan eksistensi bisnis guna menghadapi kompetisi dunia usaha yang begitu ketat. Tanpa pengelolaan dan pemanfaatan data dan informasi tersebut secara maksimal maka boleh dikatakan perusahaan tersebut tidak memaksimalkan salah satu potensi yang dimiliki untuk menghadapi persaingan.

Selain itu untuk menghadapi perkembangan dan pertumbuhan bisnis yang cepat, data dan informasi tersebut perlu dipelihara untuk kemudahan akses jika sewaktu-waktu dibutuhkan misalnya untuk keperluan analisa *customer*, transaksi penjualan dan lain sebagainya. Kemudahan akses data-data

operasional yang bersifat *historical* dapat dikembangkan menjadi informasi guna kebutuhan perencanaan atau kebutuhan strategis untuk melihat peluang ke depan dengan tujuan membuat terobosan-terobosan bisnis baru yang sesuai dengan inti bisnis yang dimiliki perusahaan.

Untuk mendukung pengelolaan data yang baik, utamanya bagi perusahaan yang memiliki data dengan *volume* besar dan tersebar ke dalam berbagai aplikasi terpisah, dewasa ini telah ada teknologi gudang data yang bisa menggabungkan data dari berbagai macam sumber data operasional dan sinkronisasi datanya dapat dilakukan secara periodik maupun *real time*, disesuaikan dengan kebutuhan yang ada. Penerapan *data warehouse* yang baik dan sejalan bersamaan dengan kebutuhan secara otomatis memudahkan pengelolaan data dan proses pengambilan informasi dari sumber data, selanjutnya adalah bagaimana data tersebut dapat disajikan secara informatif dan dapat menjawab kebutuhan dan proses bisnis, baik yang telah terjadi, sedang terjadi maupun apa yang akan terjadi di masa datang berdasarkan informasi yang disajikan.

Pentaho Data Integration

Pentaho Data Integration (PDI, juga disebut Kettle), adalah salah satu pemimpin alat integrasi data, dapat digunakan secara luas untuk semua jenis manipulasi data, seperti memigrasikan data antar aplikasi atau basis data, mengeksport data dari basis data ke dalam bentuk file, pembersihan data, dan banyak lagi (Pulvirenti & Roldán, 2011).

Data Warehouse

Data Warehouse adalah suatu sistem yang digunakan untuk melakukan ekstraksi, *cleansing*, *adjustment*, dan memberikan *source* data ke dalam penyimpanan data dimensi dan kemudian digunakan untuk mendukung dan mengimplementasikan pertanyaan dan analisa untuk tujuan *decision making* (Ralph Kimball & Caserta, 2012).

Penelitian Terkait

Adapun penelitian yang telah dipublikasikan dalam jurnal berkaitan dengan Data Warehouse diantaranya :

1. Penelitian dengan judul *Development of Data Warehouse to Improve Services in IT*

Services Company Oleh Al Faris dkk, tujuan dari penelitian ini adalah mengembangkan data warehouse untuk membantu perusahaan menjaga waktu layanan pengiriman selalu tepat sasaran. Penelitian ini dimulai dengan mengidentifikasi masalah, pengumpulan data, studi literatur, kemudian mengembangkan data warehouse, membuat analisis dengan OLAP, membuat laporan dan dashboard sesuai kebutuhan, dan akhirnya evaluasi data warehouse sudah dikembangkan. Proses Extract Transform Load (ETL) mentransfer data dari basis data operasional ke gudang data. Fitur analisis OLAP di gudang data dapat membantu perusahaan menganalisis informasi dari data multidimensi; pembuatan laporan dan dasbor dari informasi di data warehouse dapat membantu tim penjualan untuk mengetahui strategi terbaik dan pengambilan keputusan. Akibatnya, data warehouse yang telah dibangun dapat membantu tim penjualan memberikan layanan tepat waktu kepada pelanggan mereka (Al Faris, Suharjo, Diana, & Nugroho, 2018).

2. Penelitian dengan judul *Analysis Students' Graduation Eligibility Using Data Warehouse* oleh E. Chandra dkk, tujuan dari penelitian ini adalah untuk membuat sistem menggunakan data warehouse yang dapat melihat kemajuan kinerja siswa dan kursus gagal dalam setiap istilah. Sistem ini diharapkan mampu memprediksi kelulusan siswa. Metode penelitian dari penelitian ini terdiri dari beberapa langkah untuk mengimplementasikan data warehouse dan kemudian diikuti dengan menghasilkan laporan dan analisis on-line analytical processing (OLAP). Evaluasi dilakukan dengan membandingkan sistem saat ini dengan sistem yang diusulkan. Hasilnya menunjukkan sistem dapat memberikan data untuk kelayakan kelulusan (Chandra, Girsang, Hadinata, & Isa, 2018).
3. Penelitian dengan judul *Knowledge Management Study in Data Warehouse* oleh A. Gunawan dan S. Kurnia, dalam penelitian ini, modelnya adalah PT Autochem Industry, organisasi ritel HealthCare. Organisasi ini memiliki 24 cabang di Kota-kota Besar di Indonesia. Masalah muncul ketika 150 pemasar perlu menangani lebih dari 100 produk. Pemasaran mengalami beberapa kesulitan dalam memahami produk dengan

- keunikannya, dan strategi pemasaran produk perawatan kesehatan. Dengan menggunakan Knowledge Management, pemasaran dapat lebih efisien dan produktif dalam menangani pasar untuk produk perawatan kesehatan (Gunawan & Kurnia, 2018).
4. Penelitian dengan judul *Building a Data Warehouse to Support Active Student Management: Analysis and Design* oleh I. Sutedja dkk, tujuan dari penelitian ini adalah untuk menganalisis dan merancang gudang data untuk mengintegrasikan berbagai database operasional yang diperlukan untuk memberikan informasi tentang siswa aktif di Universitas XYZ. Metode analisis dilakukan dengan menjalankan analisis sistem dan analisis kelemahan sistem yang sedang berjalan. Sedangkan metode perancangan data warehouse menggunakan 4 tahap (Four-Step Metodologi) yang digunakan oleh Ralph Kimball dalam mendesain data warehouse. Tahapannya adalah memilih proses bisnis, mendeklarasikan gandum, mengidentifikasi dimensi, dan mengidentifikasi fakta. Hasil yang dicapai adalah desain data warehouse dan dashboard yang akan memberikan informasi yang relevan dan terintegrasi tentang siswa aktif yang dapat dilihat dari berbagai sudut. Gudang data yang dirancang diperlukan untuk membantu organisasi menganalisis informasi seperlunya dan membantu manajemen membuat keputusan strategis. Kesimpulannya adalah bahwa dengan membangun gudang data untuk mendukung manajemen siswa aktif dapat membantu universitas untuk menganalisis siswa aktif dan membuat keputusan di bidang siswa (Sutedja, Yudha, Khotimah, & Vasthi, 2018).
 5. Penelitian dengan judul *Design and Implementation Data Warehouse in Insurance Company* oleh R. Ari Setyawan dkk, tujuan dari penelitian ini adalah untuk mengembangkan teknik untuk menganalisis data ini dari perusahaan. Metode yang digunakan untuk mengimplementasikan analisis data dalam makalah ini adalah merancang gudang data. Menggunakan metodologi ninestep sebagai metode untuk mendesain data warehouse dan Pentaho sebagai alat untuk ETL (Extract, Transform, Load), analisis OLAP dan data pelaporan, hasil penelitian ini menyimpulkan bahwa implementasi data warehouse untuk melakukan analisis data dan pelaporan lebih baik dan lebih efektif (Ari Setyawan, Prasetyo, & Girsang, 2019).
 6. Penelitian dengan judul *A Data Warehouse Based Modelling Technique for Stock Market Analysis* oleh D. Mondal dkk, tujuan dari makalah ini adalah mengidentifikasi model gudang untuk membangun kerangka kerja analitis dan menganalisis berbagai parameter penting yang secara langsung berdampak pada perubahan pasar saham. Kami mengidentifikasi parameter yang mewakili jendela tampilan dan perspektif yang berbeda terhadap kinerja pasar saham dan tren pergerakan. Kami mengategorikan dan mendefinisikan banyak faktor intrinsik maupun eksternal yang dapat memengaruhi pasar saham secara keseluruhan. Sensex dan Nifty digunakan sebagai denyut nadi pasar saham India. Dalam makalah ini, kami fokus pada mendefinisikan model OLAP yang cocok yang dapat memenuhi semua parameter yang mempengaruhi pasar saham. Kami juga mengidentifikasi berbagai aplikasi model analitik ini untuk meramalkan informasi untuk membantu pengambilan keputusan (Mondal, Maji, Goto, C. Debnath, & Sen, 2018).
 7. Penelitian dengan judul *Designing and Implementing Data Warehouse for Agricultural Big Data* oleh V. M. Ngo dkk, dalam Penelitian ini, penulis merancang dan mengimplementasikan data warehouse pertanian tingkat kontinental dengan menggabungkan Hive, MongoDB dan Cassandra. Kemampuan data warehouse yaitu: (1) skema fleksibel; (2) integrasi data dari multi dataset pertanian; (3) ilmu data dan dukungan intelijen bisnis; (4) kinerja tinggi; (5) penyimpanan tinggi; (6) keamanan; (7) tata kelola dan pemantauan; (8) konsistensi, ketersediaan dan toleransi partisi; (9) penyebaran dan penyebaran cloud. (M.Ngo, Le-Khac, & Kechadi, 2019)
 8. Penelitian dengan judul *Nuts and Bolts of ETL in Data Warehouse* oleh S. Sachin dkk, penelitian ini memberikan pemahaman tingkat tinggi tentang mekanisme pemrosesan data warehouse yang ada termasuk pemrosesan konvensional dan pemrosesan terdistribusi. Proses Transformasi dan Pemuatan Ekstraksi yang ada akan dianalisis untuk pemahaman yang

lebih baik tentang sub proses proses pembangunan data warehouse. (Sachhin, Goyal, Avinash, & Kamal, 2019).

9. Penelitian dengan judul *Development and application of a high throughput natural language processing architecture to convert all clinical documents in a clinical data warehouse into standardized medical vocabularies* oleh M. Afshar dkk, penelitian bertujuan untuk mengembangkan arsitektur NLP menggunakan Analisis Teks klinis dan Sistem Ekstraksi Pengetahuan dan menyajikan kasus penggunaan model prediktif; Bahan dan Metode: CDW terdiri dari 1.103.038 pasien dalam 10 tahun. Arsitektur dibangun menggunakan repositori data Hadoop untuk sumber data dan 3 server pemrosesan simetris skala besar untuk NLP. Setiap entitas yang disebutkan dalam dokumen klinis dipetakan ke konsep pengidentifikasi unik Sistem Bahasa Kedokteran. (Afshar et al., 2019)
10. Penelitian dengan judul *The Implementation of a Data-Accessing Platform Built from Big Data Warehouse of Electric Loads* oleh S. Chou dkk, penelitian ini membahas tentang kontribusi kernel dari pekerjaan penelitian yang disimpan dalam gudang data dilihat dari dua bagian: (1) Modul perangkat lunak multi-layer diadopsi untuk merancang dan mengimplementasikan platform pemantauan daya real-time yang tertanam dengan beberapa karakteristik yang sangat baik dari efisiensi tinggi, kelayakan tinggi dan biaya rendah . (2) Eksperimen dasar dilakukan untuk memverifikasi efisiensi respon-respons, dan evaluasi kinerja. (Chou, Yang, Jiang, & Chang, 2018)
11. Penelitian dengan judul *Analisis Dan Perancangan Data Warehouse Perpustakaan (Studi Kasus: Perpustakaan Universitas Binadarma Palembang)* oleh Andri dan Baibul Tujni, penelitian ini akan melakukan perancangan Data Warehouse. Dalam penelitian ini akan dirancang sebuah Data Warehouse sebagai repository perpustakaan yang akan diterapkan dengan menggunakan *software Pentaho Kettle*. Hasil dari artikel ini yaitu sebuah rancangan gudang data yang digunakan sebagai repositori data-data perpustakaan. (Andri & Tujini, 2015)

METODE PENELITIAN

Metode Pengumpulan Data

Metode perancangan gudang data yang digunakan dalam penelitian ini menggunakan metode *Nine-Step Design Methodology* (Ralph Kimball, Margy Ross, Becker, Mundy, & Warren Thorntwaite, 2010). Objek dari penelitian ini adalah PT. Subafood Pangan Jaya. Metode pengumpulan data yang akan digunakan untuk penelitian ini adalah sebagai berikut :

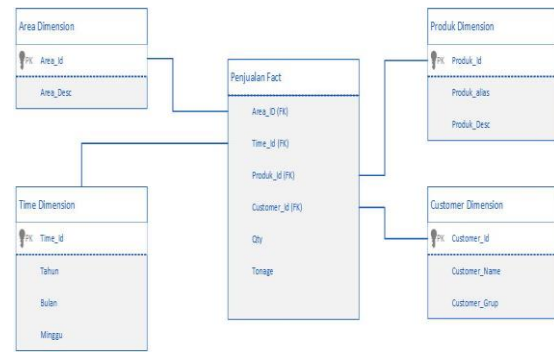
1. Pengamatan (Observasi)
Pada proses ini peneliti melakukan pengamatan secara langsung pada objek penelitian dalam hal ini adalah sistem informasi penjualan pada PT. Subafood Pangan Jaya.
2. Wawancara (Interview)
Wawancara yang peneliti lakukan untuk mendapatkan informasi mengenai proses bisnis yang ada dalam sistem informasi penjualan PT. Subafood Pangan Jaya
3. Studi Pustaka
Dalam penelitian ini peneliti juga menggunakan beberapa referensi yang digunakan sebagai sumber acuan penggunaan teori dalam penelitian ini.

HASIL DAN PEMBAHASAN

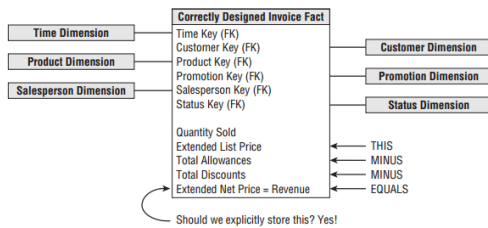
Perancangan Data Warehouse

1. Memilih Proses (*choose the process*)
Yang dimaksud dengan memilih proses adalah aktivitas operasional tertentu. Yaitu area subjek proses bisnis pertama yang ingin coba dibangun haruslah yang paling mendasar. Proses ini harus secara bersamaan menjawab pertanyaan bisnis yang paling penting dan menjadi yang paling mudah diakses dari sudut pandang ekstraksi data.
2. Memilih Grain (*choose the grain*)
Choose the grain berarti memutuskan persis apa yang diwakili oleh catatan tabel fakta. Hal yang perlu diingat adalah bahwa tabel fakta adalah pusat besar tabel dalam desain dimensi yang memiliki kunci multi bagian. Masing-masing komponen kunci *multipart* adalah kunci untuk tabel dimensi individu.
3. Mengidentifikasi dan Penyesuaian Dimensi (*Identify and conform the dimensions*)
Dimensi adalah sumber dari *query constraints* dan *row headers* yang akan digunakan dalam laporan pengguna; Seperangkat dimensi yang dirancang dengan baik membuat model dapat dimengerti dan mudah digunakan.

- Memilih Fakta (*choose the facts*)
 Butir tabel fakta menentukan fakta mana yang dapat kita gunakan untuk proses bisnis. Semua fakta harus dinyatakan pada tingkat yang seragam.
- Menyimpan *pre-Calculation* dalam Tabel Fakta (*store precalculations in the fact table*)
 Contoh umum dari kebutuhan untuk menyimpan *pre-Calculation* terjadi ketika fakta terdiri dari laporan laba rugi. Hal ini akan sering terjadi ketika tabel fakta didasarkan pada tagihan pelanggan.



Gambar 2. Fact Table Penjualan



Gambar 1. Ilustrasi Perhitungan Awal pada Tabel Fakta (Ralph Kimball et al., 2010)

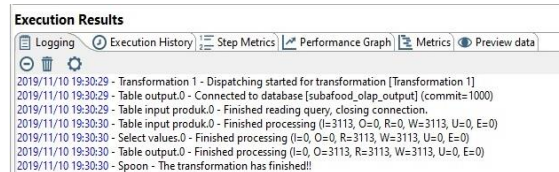
- Melengkapi Tabel Dimensi (*round out the dimension tables*)
 Pada tahapan ini tabel fakta sudah selesai, dan kita mengerti peran tabel dimensi dalam menyediakan titik masuk ke tabel fakta melalui kendala pada atribut dimensi.
- Pemilihan Durasi Database (*choose the duration of the database*)
choose the duration of the database mempunyai maksud untuk mengukur seberapa jauh ke masa lalu fakta tabel yang telah dibuat. Kebutuhan ini biasanya memberi mandat setidaknya lima perempat kalender data.

Implementasi Pentaho Data Integration (PDI)

Proses ETL mengekstraksi data dari basis data aplikasi Sistem Informasi Penjualan sebagai sumber data, mentransformasikannya, dan kemudian memuatnya ke database dalam *data warehouse* sebagai tujuan. *Pentaho Data Integration (PDI)* digunakan untuk memproses ETL. Proses ETL dalam penelitian ini bisa dilihat pada gambar dibawah ini:



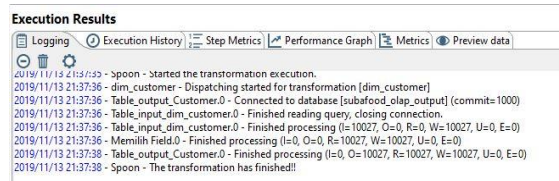
Gambar 3. Proses ETL dimensi Produk



Gambar 4. Rincian Hasil Proses ETL dimensi Produk



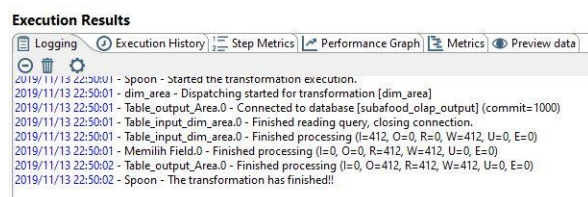
Gambar 5. Proses ETL dimensi Customer



Gambar 6. Rincian Hasil Proses ETL dimensi Customer



Gambar 7. Proses ETL dimensi Area Sales



Gambar 8. Rincian Hasil Proses ETL dimensi Area Sales

SIMPULAN DAN SARAN

Kesimpulan dari artikel ini adalah hasil dari perancangan gudang data yang telah dirancang dapat membantu pimpinan penjualan dalam pengambilan keputusan kaitanya dengan pembuatan strategi penjualan. *Data Warehouse* yang telah dibangun dapat dipergunakan oleh pihak yang mempunyai kepentingan untuk dapat digunakan sebagai sumber data dalam mengelola dan mengolah data penjualan.

Saran dari artikel ini adalah hasil dari perancangan gudang data yang telah dirancang dapat digunakan untuk membuat aplikasi *Business Intelligence* agar bisa mendapatkan *insight* lebih dalam mengenai penjualan

DAFTAR PUSTAKA

- Afshar, M., Dligach, D., Sharma, B., Cai, X., Boyda, J., Birch, S., ... Price, R. (2019). Development and application of a high throughput natural language processing architecture to convert all clinical documents in a clinical data warehouse into standardized medical vocabularies. *Journal of the American Medical Informatics Association*, *0*(0), 1–6. <https://doi.org/10.1093/jamia/ocz068>
- Al Faris, F. Z., Suharjo, Diana, & Nugroho, A. (2018). Development of Data Warehouse to Improve Services in IT Services Company. *Proceedings of 2018 International Conference on Information Management and Technology, ICIMTech 2018*, (September), 483–488. <https://doi.org/10.1109/ICIMTech.2018.8528146>
- Andri, & Tujini, B. (2015). Analisis Dan Perancangan Data Warehouse Perpustakaan (Studi Kasus: Perpustakaan Universitas Binadarma Palembang). *Seminar Nasional Informatika, 2015*(November), 28–34.
- Ari Setyawan, R., Prasetyo, E., & Girsang, A. S. (2019). Design and Implementation Data Warehouse in Insurance Company. *Journal of Physics: Conference Series*, *1175*(1). <https://doi.org/10.1088/1742-6596/1175/1/012072>
- Chandra, E., Girsang, A. S., Hadinata, R., & Isa, S. M. (2018). Analysis Students' Graduation Eligibility Using Data Warehouse. *Proceedings of 2018 International Conference on Information Management and Technology, ICIMTech 2018*, (September), 61–64. <https://doi.org/10.1109/ICIMTech.2018.8528119>
- Chou, S. C., Yang, C. T., Jiang, F. C., & Chang, C. H. (2018). The Implementation of a Data-Accessing Platform Built from Big Data Warehouse of Electric Loads. *Proceedings - International Computer Software and Applications Conference*, *2*, 87–92. <https://doi.org/10.1109/COMPSAC.2018.10208>
- Gunawan, A., & Kurnia, S. G. (2018). Knowledge Management Study in Data Warehouse. *Proceedings of 2018 International Conference on Information Management and Technology, ICIMTech 2018*, (September), 1–5. <https://doi.org/10.1109/ICIMTech.2018.8528121>
- M.Ngo, V., Le-Khac, N.-A., & Kechadi, M.-T. (2019). Designing and Implementing Data Warehouse for Agricultural Big Data. *Big Data*, *11514*, 18–32. <https://doi.org/10.1007/978-3-030-23551-2>
- Mondal, D., Maji, G., Goto, T., C. Debnath, N., & Sen, S. (2018). A Data Warehouse Based Modelling Technique for Stock Market Analysis. *International Journal of Engineering & Technology*, *7*(3.13), 165. <https://doi.org/10.14419/ijet.v7i3.13.17325>
- Pulvirenti, A. S., & Roldán, M. C. (2011). Pentaho Data Integration 4 Cookbook. In *Integration The Vlsi Journal*.
- Ralph Kimball, & Caserta, J. (2012). The Data Warehouse ETL Toolkit. In *Wiley* (Vol. 66).
- Ralph Kimball, Margy Ross, Becker, B., Mundy, J., & Warren Thornthwaite. (2010). *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence*. Wiley Publishing, Inc.
- Sachhin, S., Goyal, S. K., Avinash, S., & Kamal, K. (2019). Nuts and Bolts of ETL in Data Warehouse. In *Emerging Trends in Expert Applications and Security* (Vol. 841). <https://doi.org/10.1007/978-981-13-2285-3>
- Sutedja, I., Yudha, P., Khotimah, N., & Vasthi, C. (2018). Building a Data Warehouse to Support Active Student Management: Analysis and Design. *Proceedings of 2018 International Conference on Information Management and Technology, ICIMTech 2018*, (September), 460–465. <https://doi.org/10.1109/ICIMTech.2018.8528196>