NIM : 512121230007

NAMA : TEGUH AGUNG PRABOWO

JURUSAN : SISTEM INFORMASI

MATAKULIAH : SISTEM PENUNJANG KEPUTUSAN

1. **Data Mining**
2. Pengertian Data Mining

Data mining adalah suatu proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar. Proses data mining seringkali menggunakan metode statistika, matematika, hingga memanfaatkan teknologi artificial intelligence.

Nama alternatifnya yaitu Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, dan lain-lain.

Dalam proses KDD tersebut termasuk melakukan pembersihan data, integrasi data, seleksi data, transformasi, penambangan data, evaluasi pola, dan presentasi pengetahuan.

1. Fungsi Data Mining

Data mining memiliki banyak sekali fungsi, Untuk fungsi utamanya sendiri yaitu ada dua; Yaitu fungsi descriptive dan fungsi predictive.

1. Descriptive

Fungsi deskripsi dalam data mining adalah sebuah fungsi untuk memahami lebih jauh tentang data yang diamati. Dengan melakukan sebuah proses diharap bisa mengetahui perilaku dari sebuah data tersebut. Data tersebut itulah yang nantinya dapat digunakan untuk mengetahui karakteristik dari data yang dimaksud.

Dengan menggunakan Fungsi descriptive Data mining, Maka nantinya bisa menemukan pola tertentu yang tersembunyi dalam sebuah data. Dengan kata lain jika pola yang berulang dan bernilai itulah karakteristik sebuah data bisa diketahui.

1. Predictive

Fungsi prediksi merupakan sebuah fungsi bagaimana sebuah proses nantinya akan menemukan pola tertentu dari suatu data. Pola-pola tersebut dapat diketahui dari berbagai variabel-variabel yang ada pada data.

Ketika sudah menemukan pola, Maka pola yang didapat tersebut bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya.

Karena itulah fungsi satu ini dikatakan sebagai fungsi prediksi sama halnya dengan melakukan predictive analisis. Fungsi ini juga bisa digunakan untuk memprediksi sebuah variabel tertentu yang tidak ada dalam suatu data.

Sehingga fungsi ini memudahkan dan menguntungkan bagi siapapun yang memerlukan prediksi yang akurat untuk membuat hal penting tersebut menjadi lebih baik.

Fungsi Data mining yang lainnya yaitu : characterization, discrimination, association, classification, clustering, outlier and trend analysis.

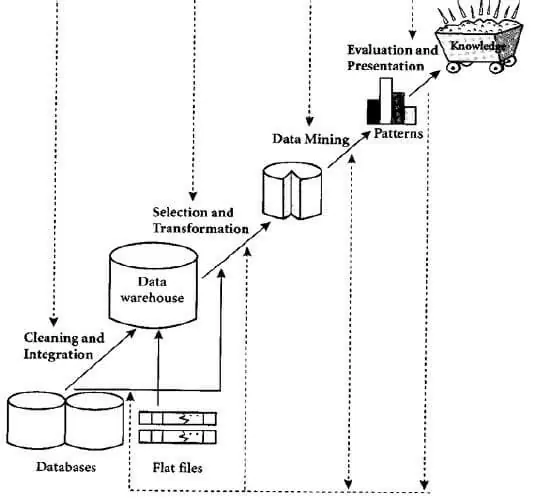
* Multidimensional concept description, Karakterisasi dan diskriminasi, Atau berfungsi untuk Menggeneralisasikan, meringkas, dan membedakan karakteristik data, dll.
* Frequent patterns, association, correlation
* Classification and prediction, Membangun model (fungsi) yang menggambarkan dan membedakan kelas atau konsep untuk prediksi masa depan. Misalnya, Mengklasifikasikan negara berdasarkan (iklim), atau mengklasifikasikan mobil berdasarkan (jarak tempuh gas)
* Cluster analysis, Membuat data grup untuk membentuk kelas baru. Misalnya, Memaksimalkan kesamaan intra-kelas & meminimalkan kesamaan antar kelas
* Outlier analysis, Objek data yang tidak sesuai dengan perilaku umum dari data, Berguna dalam deteksi penipuan, analisis peristiwa langka.
* Trend and evolution analysis, Trend dan penyimpangan: misalnya Analisis regresi atau Mining Penambangan pola berurutan: misalnya, Kamera digital, atau Analisis periodisitas dan Analisis berbasis kesamaan.
* Other pattern-directed or statistical analyses

1. Metode Data Mining

Dalam melakukan pengumpulan informasi tentunya terdapat metode, Metode tersebut akan membantu dalam proses menemukan data. Data mining akan menyediakan perencanaan dari ide hingga implementasi akhir.

1. Proses pengambilan Data

Bagaimana proses pengambilan data? Tadi di atas sudah dijelaskan tentang KDD atau Knowledge discovery (mining) in databases. Dengan KDD tersebut itulah kamu dapat melakukan proses pengambilan data.



Proses atau tahapan-tahapan tersebut dimulai dari data mentah dan berakhir dengan pengetahuan atau informasi yang telah diolah. Nah proses tersebut sebagai berikut:

1. Data Cleansing, Proses dimana data-data yang tidak lengkap, mengandung error dan tidak konsisten dibuang dari koleksi data. Ketahui juga data lifecycle management untuk mengetahui tentang pengolahan data.
2. Data Integration, Proses integrasi data dimana yang berulang akan dikombinasikan.
3. Selection, Proses seleksi atau pemilihan data yang relevan terhadap analisis untuk diterima dari koleksi data yang ada.
4. Data Transformation, Proses transformasi data yang sudah dipilih ke dalam bentuk mining procedure melalui cara dan agresi data.
5. Data Mining, Proses yang paling penting dimana akan dilakukan berbagai teknik yang diaplikasikan untuk mengekstrak berbagai pola-pola potensial untuk mendapatkan data yang berguna.
6. Pattern Evolution, Sebuah proses dimana pola-pola menarik yang sebelumnya sudah ditemukan dengan identifikasi berdasarkan measure yang telah diberikan
7. Knowledge Presentation, Merupakan proses tahap terakhir, Dalam hal ini digunakan teknik visualisasi yang bertujuan membantu user dalam mengerti dan menginterpretasikan hasil dari penambangan data.
8. Teknik dalam Proses Penambangan Data

Terdapat berbagai macam teknik yang digunakan dalam proses penambangan data, Apa saja teknik yang bisa digunakan dalam proses Data Mining?

1. Predictive Modeling, Terdapat dua teknik yaitu Classification dan Value Prediction
2. Database Segmentation, Melakukan partisi database menjadi sejumlah segmen, cluster, atau record yang sama
3. Link analysis, Sebuah teknik untuk membuat hubungan antara record yang individu atau sekumpulan record dalam database.
4. Deviation detection, Sebuah teknik untuk mengidentifikasi outlier yang mengekspresikan sebuah deviasi dari ekspektasi yang sudah diketahui sebelumnya.
5. Nearest Neighbour, Yaitu teknik yang memprediksi pengelompokan, Teknik ini sendiri merupakan teknik yang tertua yang digunakan dalam data mining.
6. Clustering, merupakan teknik untuk mengklasifikasikan data berdasarkan kriteria masing-masing data.
7. Decision Tree, Merupakan teknik generasi selanjutnya, dimana teknik ini adalah sebuah model prediktif yang dapat digambarkan seperti pohon. Setiap node yang terdapat dalam struktur pohon tersebut mewakili sebuah pertanyaan yang digunakan untuk menggolongkan data.
8. **Datalake**

Data lake merupakan pusat berkumpulnya data-data dalam format dan skala aslinya. Kamu bisa menyimpan berbagai jenis data tanpa perlu menyusunnya dalam struktur, pengelompokan, atau hierarki tertentu. Dengan kata lain, data yang terdapat dalam data lake adalah data mentah yang belum diproses atau dianalisis. Data lake dapat digunakan untuk menyimpan data dari beragam sumber. Data-data di dalamnya pun terdiri dari berbagai tipe dan skema. Berbagai macam pengguna dari mana saja dapat mengakses data lake dan mengambil sampel data dari dalamnya.

Terdapat sejumlah komponen yang menyusun data lake, yakni:

1. Data Ingestion and Storage yang berguna untuk menerima data, baik secara real time atau dalam kelompok. Komponen ini juga memungkinkan pengguna untuk menyimpan dan mengakses data.
2. Data Processing, yaitu kemampuan untuk bekerja dengan data mentah sehingga dapat dianalisis melalui proses standar.
3. Data Analysis yang merupakan modul-modul dengan fungsi memperoleh hasil analisis sistematis dari suatu data
4. Data Integration, atau kemampuan untuk menghubungkan aplikasi dengan platform. Namun, pertama-tama data harus diekstraksi terlebih dulu dalam format sesuai kebutuhan

Komponen-komponen pada data lake menyediakan berbagai fungsi yang membantu perusahaan untuk mendapatkan lebih banyak konsumen, meningkatkan produktivitas, dan membuat keputusan.

Semuanya berkontribusi untuk meningkatkan pertumbuhan bisnis dengan pesat. Keuntungan tersebut dapat kamu peroleh sebagai berikut:

1. Mengindeks data

Kita dapat menyimpan berbagai jenis data dan database. Termasuk di antaranya data operasional, data dari aplikasi bisnis, atau data yang bersifat non-relasional seperti data yang diperoleh dari aplikasi mobile dan media sosial. Meskipun ini merupakan data mentah, kita bisa memahami isi data dengan adanya katalog, crawling, dan indeks data.

1. Machine learning

Perusahaan dapat memperoleh gambaran operasional dan marketing melalui data yang diperoleh dari data lake. Data-data ini menggambarkan tren serta pola perilaku konsumen. Kemudian, perusahaan dapat menerapkan machine learning untuk membuat model prediksi dan perkiraan dari data-data tersebut.

1. Mengembangkan interaksi dengan konsumen

Data lake mampu menggabungkan data konsumen dari platform CRM dengan hasil analisis media social. Penggabungan tersebut juga dapat dilakukan dengan platform marketing yang menggambarkan riwayat pembelian konsumen. Hal ini berguna agar perusahaan dapat mengidentifikasi mana konsumen yang paling menguntungkan, apa yang melatarbelakangi pola perilaku konsumen, serta reward seperti apa yang dapat meningkatkan kesetiaan konsumen.

1. Analisis

Keberadaan data lake memungkinkan para Data Scientist, pengembang data, serta siapa pun yang bergelut dalam bidang terkait untuk mengakses data sesuai kerangka dan perangkat analisis yang mereka miliki. Kamu juga dapat melakukan analisis tanpa perlu memindahkan data dari satu sistem ke sistem yang lain.

1. Murah

Data lake ringan di kantong untuk diimplementasikan. Sebab, sebagian besar teknologi untuk mengelolanya yakni open source (Hadoop) dan dapat diinstal pada perangkat keras berbiaya rendah.

Semakin banyak perusahaan mulai menggunakan data lake guna memperoleh informasi mengenai bisnis dan konsumen yang dapat diakses dengan mudah. Meski demikian, perusahaan tetap perlu menetapkan model sistem, proses, dan pengaturan agar manfaat yang diberikannya lebih optimal.

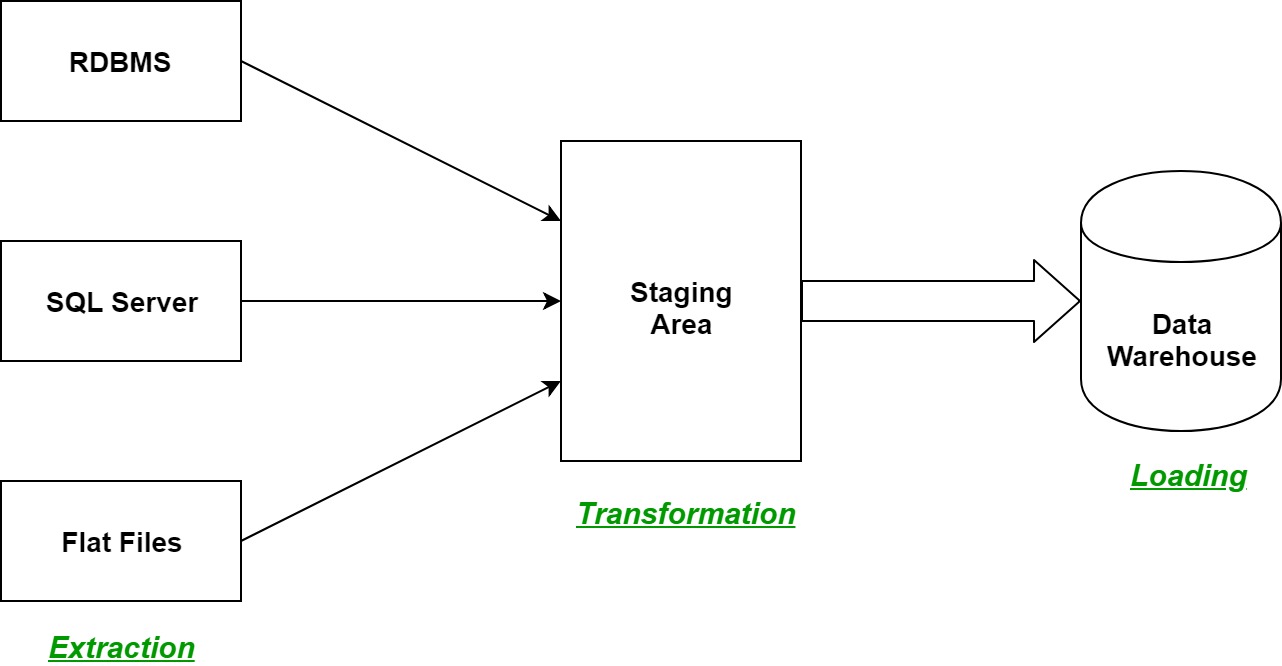
1. **ETL (Ekstrak, Transform, and Load)**
   1. Pengertian

ETL adalah singkatan dari extract, transform, dan load. Melansir IBM, ia merupakan proses integrasi data dan akan dikombinasikan dari berbagai sumber. Setelah itu, mereka disimpan di tempat bernama data warehouse. Dalam proses ini, kata Guru99, ada tools yang bisa kamu gunakan. Misalnya, MarkLogic, Oracle, Amazon RedShift, dan lain-lain. Di perusahaan, orang yang bertanggung jawab atasnya merupakan ETL developer.

* 1. Manfaat

ETL ternyata sangat berkaitan dengan penerapan data science dalam bisnis. Ingat, di masa kini, big data adalah unsur yang penting untuk perusahaan. Berbagai pihak sangat membutuhkannya. ETL adalah salah satu proses yang bisa membantu mereka semua. Lewatnya, berbagai informasi bisa diambil dan dimanfaatkan. Dengan begitu, keputusan bisnis yang tepat pun bisa dibuat. Tak ada lagi pihak yang meraba-raba dalam kegelapan atau “asal tembak”. Terlebih lagi, ETL memungkinkan perusahaan melakukan data governance. Dengannya, informasi pun terkumpul menjadi satu, tak tercecer ke mana-mana.

* 1. Proses ETL



Source: geeksforgeeks.org

Dirangkum dari Geeks for Geeks, penjelasan langkah dalam ETL adalah:

* + 1. Extraction

Langkah pertama bernama extraction. Layaknya namanya, dalam proses ini, kita akan mengambil data dari berbagai sumber. Misalnya, dari server SQL, XML, atau flat files. Nah, setelah diambil, kita tak serta-merta menaruhnya di warehouse. Tempat untuk data ini adalah staging area.

Mungkin, kita akan bertanya apa pentingnya langkah itu dalam ETL? Bukankah lebih cepat jika data langsung pergi ke tempat seharusnya?

Format dari data tersebut berbeda-beda. Belum lagi, ada kemungkinan informasi tersebut bersifat corrupt.

Oleh karena itu, kita harus menaruhnya di staging area dulu. Di sana, langkah selanjutnya pun dilakukan.

* + 1. Transformation

Tahap ETL selanjutnya adalah transformation. Pada langkah ini, data akan diolah sehingga punya satu format yang sama.

Biasanya, ada 5 hal yang dilakukan pada data:

* filtering, menyaring data dengan filter tertentu
* cleaning, menyesuaikan format penulisan, misalnya “Amerika Serikat” diubah jadi “AS”
* joining, ciri data yang serupa menjadi satu
* splitting, memecah ciri data yang berbeda menjadi dua atau lebih
* sorting mengurutkan data berdasarkan ciri tertentu

Dengan begitu, informasi yang didapatkan lebih rapi. Mereka pun siap menuju tahap selanjutnya.

* + 1. Loading

Langkah terakhirnya bernama loading. Akhirnya, data yang selesai diproses masuk ke data warehouse. Kadang kala, proses ini terjadi sangat cepat. Tiap data selesai diolah, ia langsung menjalani proses loading. Akan tetapi, kita bisa mengatur alirannya menjadi beberapa saat sekali. Dengan begitu, tingkat keseringannya bisa kamu atur.

1. **Data Warehouse**
   1. Pengertian Data Warehouse

Definisi umum dari gudang data atau data warehouse adalah sistem komputer yang bertugas untuk mengarsipkan dan menganalisis sebuah data historis untuk keperluan suatu organisasi atau bisnis tertentu. Informasi yang dikelola dapat berupa data terkait penjualan, gaji, serta informasi harian lainnya.

Dengan menganalisa sebuah data secara terstruktur, maka dapat menghasilkan informasi yang lebih akurat untuk menunjang pengambilan keputusan oleh suatu perusahaan. Itulah mengapa, gudang data termasuk dalam salah satu parameter penunjang kegiatan business intelligence.

Gudang data sendiri pertama kali digunakan dan ditemukan pada tahun 1980-an yang saat itu juga banyak bisnis besar yang menerapkan DSS (Decision Support System). Hingga saat ini, banyak perusahaan yang menerapkan strategi tersebut untuk membantu dalam meningkatkan kualitas dan produktivitas kerja.

* 1. Fungsi Data Warehouse
     1. Membantu dalam Mengambil Keputusan

Fungsi pertama dari data warehouse adalah untuk dapat mengambil sistem keputusan yang tepat. Dimana, informasi tersebut haruslah kredibel dan berdasarkan fakta yang ada. Tujuannya ialah agar dapat membuat keputusan yang benar dalam segmentasi pasar, manajemen inventaris, dan manajemen keuangan (finansial).

* + 1. Mudah dan Cepat dalam Mengakses Data

Fungsi yang kedua adalah mudah untuk diakses, dimana pengguna dapat berinteraksi dengan sumber secara cepat dan lebih efisien. Yang mana, keefektifan tersebut dapat memberikan impact pada pengambilan keputusan yang lebih cepat daripada kompetitor bisnis anda.

* + 1. Menciptakan Konsistensi Data

Dikarenakan data warehouse merupakan sistem yang dapat mengumpulkan berbagai informasi dari platform yang berbeda, sehingga dapat diubah menjadi format tunggal agar dapat digunakan dengan baik.

Kemudian, untuk kedepannya perusahaan anda dapat memperoleh hasil yang maksimal dan konsisten satu sama lainnya. Ketika standarisasi data telah terbentuk, maka anda dapat menaruh kepercayaan yang tinggi pada tingkat keakuratan gudang data untuk dapat membuat keputusan bisnis yang jelas dan tepat.

* + 1. Menghasilkan ROI yang Optimal

ROI (Return of Investment) adalah tingkat rasio antara jumlah laba bersih dan biaya investasi yang dihasilkan dari beberapa investasi pada resource (sumber daya) yang ada. Jika dihubungkan dengan data warehouse sendiri, memiliki fungsi khusus untuk memberikan pengembalian investasi yang lebih menguntungkan.

* + 1. Menyediakan Historical Intelligence

Fungsi yang terakhir adalah sebagai penyedia mayoritas historical data agar anda dapat menganalisis setiap periode atau tren waktu yang berbeda, untuk dapat membuat prediksi bisnis (forecasting) di masa yang akan datang.

* 1. Karakteristik Data Warehouse

1. Berorientasi pada Subjek

Data yang berorientasi pada subjek memiliki dimensi waktu yang saling terintegrasi. Dimana, nantinya dapat mendukung proses pengambilan keputusan yang dilakukan oleh para manager pada tiap jenjangnya.

1. Data telah Terintegrasi

Dikarenakan gudang data merupakan kumpulan data dari berbagai sistem informasi yang ada, maka informasi yang diperoleh juga mampu untuk terintegrasi dengan baik. Sehingga, dapat membantu percepatan pengambilan keputusan secara efektif dan efisien.

1. Data Bersifat Tetap

Karakteristik yang ketiga yaitu, data yang dipakai bersifat tetap. Maksudnya disini adalah informasi yang digunakan tidak berganti dan tidak dapat diubah secara paksa. Dikarenakan, data tersebut pada umumnya berisi kumpulan informasi yang telah ada dan terdapat jangka waktu yang telah digunakan.

1. Dibuat dalam Rentang Waktu Tertentu

Sistem basis data (database) yang digunakan tentunya memiliki rentang waktu tertentu sesuai dengan kebutuhan anda. Dimana, hal tersebut akan berpengaruh pada pemakaian untuk setiap strategi bisnis yang anda jalankan. Sehingga, basis data yang terbentuk akan selalu diikuti dengan interval waktu penggunaannya.

* 1. Komponen Data Warehouse

1. Gudang

Komponen yang pertama adalah gudang atau tempat penyimpanan dari data tersebut. Bentuknya pun bisa berbeda – beda disesuaikan dengan kebutuhan anda, diantaranya adalah meliputi data warehouse appliance, analytics, cloud – hosted, dan typical relational database.

1. Manajemen Gudang Data

Peran dari manajemen gudang data sangatlah kompleks, dimana harus ada pihak yang mampu untuk mengelola gudang data dengan baik. Hal yang harus diperhatikan dalam pengelolaan gudang data antara lain, dari segi keamanan, pembaharuan data, pemilihan prioritas tugas, serta mengelola backup dan recovery apabila terdapat suatu keadaan yang dianggap mendesak.

1. Metadata

Komponen yang ketiga adalah metadata, yaitu sebuah keterangan singkat dalam data tersebut. Fungsi dari metadata sendiri adalah memberikan konteks atau gambaran mengenai sebuah informasi agar tampak lebih jelas dan selaras.

1. Tools Akses

Terdapat beberapa tools yang dapat anda manfaatkan untuk mendukung strategi bisnis anda menggunakan data warehouse. Pertama, anda dapat menggunakan konsep data mining dan OLAP.

Untuk lebih jelasnya, anda dapat mencari beberapa referensi yang ada dalam internet. Kedua, anda dapat menggunakan query reporting atau mengembangkan tools analytics sendiri apabila anda memiliki tim developer atau pengembang aplikasi.

1. Tools ETL

ETL merupakan singkatan dari Extract, Transform, and load, yang berarti sebuah proses pengambilan data yang nantinya dapat dimodifikasi formatnya. Dan pada akhirnya, deretan informasi tersebut akan dimasukkan ke dalam data warehouse. Tools tersebut nantinya akan mempengaruhi waktu, metode, dan tipe modifikasi dari data tersebut.

* 1. Star Schema Data Warehouse

Banyak sekali skema yang dapat anda gunakan untuk memudahkan dalam memproses kebutuhan data warehouse. Salah satunya dengan menggunakan metode star schema yang cukup efektif dalam mengolah sebuah database.

Skema ini terbilang sederhana dan susunan tabelnya membentuk pola bintang yang di dalamnya tersusun atas dua entitas, yaitu tabel fakta dan dimensi. Berikut merupakan beberapa penjelasan terkait kedua istilah tersebut.

1. Tabel Fakta

Tabel fakta atau fact table berisi metrik mengenai proses bisnis sebuah perusahaan atau organisasi. Data yang termuat dalam tabel tersebut bersifat numerik (angka) dan dapat ditambah. Yang perlu diperhatikan adalah dalam data tersebut harus termuat hal – hal yang berhubungan dengan tabel di sekitarnya, yaitu tabel dimensi.

1. Tabel Dimensi

Tabel dimensi atau dimension table memuat beberapa informasi terkait kapan, di mana, apa, dan informasi yang lainnya. Berbanding terbalik dengan tabel fakta, dimension table berisi informasi yang bersifat kualitatif. Dimana, dalam metode star schema sendiri terdapat beberapa tabel dimensi sekaligus yang mengelilingi tabel fakta.

* 1. Kesimpulan

Dalam menggunakan data warehouse, pastikan anda telah menyusun dan mengelola database dengan baik serta membangun kebutuhan sistem perusahaan yang optimal. Gunakan metode yang sesuai dengan kebutuhan bisnis perusahaan anda. Kemudian, pergunakan juga tools yang dapat mendukung proses manajemen data anda dengan lebih baik dan optimal.