

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Teori Umum

Dalam penelitian ini dibutuhkan beberapa tinjauan teori umum yang mendukung literasi dalam tahapan perancangan di penelitian ini

2.1.1 Perancangan

Tahap pertama dalam pembuatan sebuah sistem adalah perancangan sistem tersebut. Perancangan merupakan langkah pengembangan spesifikasi baru berdasarkan hasil analisis sistem yang telah direkomendasikan (Nugroho 2016). Perancangan dibutuhkan untuk memberikan gambaran-gambaran secara detail dalam membangun sebuah sistem ataupun aplikasi yang bisa menjadi pondasi bagi programmer sehingga sistem yang dirancang dapat memenuhi kebutuhan yang telah ditentukan dalam kegiatan analisis sebelumnya.

Perancangan adalah suatu proses pemilihan dan pemikiran yang menghubungkan fakta-fakta berdasarkan asumsi-asumsi yang berkaitan dengan masa datang dengan menggambarkan dan merumuskan kegiatan-kegiatan tertentu yang diyakini diperlukan untuk mencapai tujuan-tujuan tertentu dan menguraikan bagaimana pencapaiannya (Cahyaningtyas and Iriyani 2015).

2.1.2 Data

Data merupakan realitas yang menggambarkan peristiwa dan entitas nyata yang terjadi pada saat tertentu dalam dunia bisnis. Data merupakan bahan mentah

untuk diolah, yang hasilnya kemudian menjadi sebuah informasi. Dimana data tersebut diolah dan diinterpretasikan menjadi sesuatu yang berarti bagi pengambilan keputusan. (Amrizal and Harman 2022).

Data merupakan deskripsi mengenai objek, peristiwa, aktivitas, dan transaksi yang memiliki makna atau tidak memiliki dampak langsung terhadap pengguna (Anggraeni and Irviani 2017).

Beberapa contoh tipe data:

- a. Data yang terformat adalah data yang telah diatur dalam suatu format khusus, seperti data yang mencatat tanggal, jam, atau nilai mata uang.
- b. Teks adalah kumpulan huruf, angka, dan simbol-simbol khusus (contohnya + dan \$) yang penyusunannya tidak bergantung pada setiap elemen secara terpisah. Artikel koran merupakan salah satu contoh teks.
- c. Citra (*image*) adalah data yang berbentuk visual seperti foto, grafik, hasil rontgen, atau tanda tangan, dan lain sebagainya..
- d. Audio adalah data dalam bentuk suara, contohnya termasuk instrumen musik, suara manusia atau hewan, gemericik air, dan detak jantung.
- e. Video merupakan data yang terdiri dari serangkaian gambar bergerak yang mungkin juga disertai dengan suara. Video digunakan untuk merekam berbagai kejadian atau aktivitas.

Dalam bukunya "*Management Informations System: Conceptual Foundations, Structures, and Development*," Gordon B. Davis menggambarkan data sebagai materi mentah bagi informasi, yang diwujudkan sebagai sekelompok simbol tidak teratur yang mencerminkan kuantitas, tindakan, atau hal-hal lainnya.

2.1.3 Perancangan Data

Perancangan data merupakan sebuah proses membuat desain data yang akan mendukung operasional dan tujuan perusahaan (Gat 2015). Diperlukan perancangan data yang cermat dan tepat guna agar data yang dibangun dapat sesuai dengan kebutuhan operasional dan mencapai tujuan perusahaan. Perancangan data menjadi aspek penting guna memastikan bahwa data yang ada dapat menggambarkan relasi yang muncul dari proses kegiatan operasional perusahaan dengan baik.

Perancangan data dilakukan untuk mengatasi berbagai masalah yang muncul selama proses perancangan data, yang antara lain adalah:

a. Redundansi dan inkonsistensi data

Jika data yang sama disimpan berulang-ulang di beberapa file, hal ini dapat menyebabkan timbulnya ketidaksesuaian atau inkonsistensi data.

b. Kesulitan dalam pengaksesan data

Ketika belum adanya program yang dirancang maka akan timbul kesulitan dimana data yang ingin dicari tidak dapat ditemukan dengan secara langsung dan cepat.

c. Isolasi data untuk standarisasi

Jika data yang dimiliki tersebar dalam bentuk format yang tidak sama, maka akan timbul kesulitan ketika dalam menulis/merancang program. Dimana program akan sulit mengakses data-data yang memiliki format yang berbeda-beda.

d. Masalah keamanan (*security issue*)

Tidak semua pengguna (*user*) diperbolehkan dalam mengakses data. Dengan adanya perancangan data, maka bisa data-data yang penting dan data yang memiliki hak akses tertentu bisa di klasifikasikan dan diatur sedemikian rupa sehingga tidak sembarangan *user* bisa mengakses data tersebut.

e. Masalah integrasi data

Setiap data yang dimiliki akan berisi data-data yang saling berkaitan antara satu data dengan data lainnya. Dengan adanya perancangan data ini, data tersebut dapat dirancang untuk saling memiliki relasi disetiap data nya sehingga informasi yang dihasilkan dapat dipercaya.

Tujuan dan manfaat utama dengan adanya perancangan data adalah agar pengguna (*user*) dapat merancang suatu bayangan dari abstraksi data. Dimana gambaran dari data-data tersebut merupakan kondisi yang dihadapi oleh pengguna sehari-hari.

Dalam kegiatan operasional bisnis, peristiwa yang sering terjadi adalah transaksi, yaitu aktivitas yang menyebabkan perubahan nilai. Untuk menghasilkan informasi yang relevan bagi pengambilan keputusan yang tepat dan akurat, diperlukan perancangan data yang memastikan data diproses dengan baik. Oleh karena itu, data mentah harus diolah dan diproses terlebih dahulu sebelum dapat digunakan sebagai informasi yang bermanfaat bagi perusahaan (Sudjiman 2018).

2.2. Tinjauan Teori Khusus

Dalam penelitian ini dibutuhkan beberapa tinjauan teori khusus yang mendukung literasi dalam tahap analisis dan perancangan di penelitian ini.

2.2.1. *Data Warehouse*

Data warehouse adalah tipe basis data yang berperan sebagai tempat penyimpanan untuk kumpulan data terkini dan data historis yang berasal dari berbagai sumber. Data dan informasi yang dikumpulkan diperoleh dari sumber internal dan eksternal, lalu digabungkan menjadi satu kumpulan data. Seluruh data ini disimpan dalam media penyimpanan dengan kapasitas yang besar. Proses penyimpanan data dilakukan melalui transformasi ke bentuk penyimpanan multidimensional yang dirancang khusus untuk keperluan querying, pelaporan, dan eksplorasi data atau data mining (Suni 2018).

Data warehouse merupakan sebuah kumpulan data yang disimpan dalam basis data dan berfungsi sebagai dukungan untuk pengambilan keputusan dalam suatu organisasi. Data dikumpulkan dari berbagai aplikasi yang ada. Setelah proses pengumpulan data, dilakukan validasi dan restrukturisasi sebelum akhirnya disimpan dalam *data warehouse*.

Data Warehouse memiliki karakteristik yang setara dengan sistem pendukung keputusan. Karakteristik utama *data warehouse* lebih berfokus pada aplikasi pendukung keputusan. Berdasarkan definisi ini, *Data warehouse* dianggap memiliki perbedaan karakteristik yang jelas dibandingkan dengan sistem pendukung operasional lainnya, terutama karena tujuan utamanya adalah menjadi alat bantu dalam pengambilan keputusan. (Hendro Poerbo Prasetya 2016).

2.2.1.1. Karakteristik *Data Warehouse*

Dengan penjelasan *data warehouse*, adapun terdapat 6 karakteristik *data warehouse* yaitu sebagai berikut:(Subuh and Yasman 2019)

1. Berorientasi pada Objek (*Subject-Oriented*)

Data warehouse berfokus pada objek (*Subject-Oriented*) yang mempresentasikan dan menyusun data berdasarkan subjek yang diperlukan dalam proses pengambilan keputusan. Data diorganisasi dalam dimensi yang mencakup periode waktu, riwayat, wilayah, dan elemen lainnya.

2. Terintegrasi (*Integrated*)

Data warehouse memiliki sifat terintegrasi (*Integrated*), yang berarti dibangun melalui penggabungan berbagai sumber data, terutama dari berbagai aplikasi database. Hal ini menyatukan berbagai data menjadi satu kesatuan yang utuh. Karakteristik terintegrasi ini merupakan hasil dari pendekatan berfokus pada subjek.

3. *Time Variant*

Data warehouse bersifat *Time Variant*, mengidentifikasi data yang dikumpulkan dari berbagai sumber berdasarkan periode waktu penyimpanannya. Informasi yang disajikan oleh data-data dalam *data warehouse* dilihat dari perspektif sejarah atau riwayat penyimpanannya.

4. *Non-Volatile*

Data warehouse bersifat *Non-Volatile*, yang berarti data yang dikumpulkan dari berbagai sumber dan disimpan dalam *data warehouse* tidak dapat diubah, seperti diedit, diperbarui, atau dihapus. *Data warehouse* menekankan

pentingnya data historis yang asli untuk mendukung analisis data di masa depan.

5. *Accessible*

Data warehouse harus mudah diakses (*Accessible*) oleh pengguna. Pengguna dapat dengan mudah memperoleh data sesuai kebutuhan mereka, baik secara keseluruhan maupun sebagian, tergantung pada hak akses yang diberikan oleh sistem atau pemilik data.

6. *Process-Oriented*

Data warehouse dipandang sebagai proses berorientasi (*Process-Oriented*) dalam pengolahan data menjadi informasi dan menyampaikan informasi tersebut. Proses menjadi fokus utama dalam operasional *data warehouse*.

2.2.1.2. Manfaat *Data Warehouse*

Data warehouse yang telah direncanakan secara cermat merupakan landasan kunci bagi kesuksesan setiap program *Business Intelligence* (BI) atau analitik. Perannya sangat penting dalam mendukung pembuatan laporan, dashboard, dan alat analisis yang menjadi kebutuhan utama dalam operasional bisnis saat ini.

Data warehouse menyediakan informasi yang relevan untuk mendukung pengambilan keputusan berdasarkan data, membantu dalam membuat keputusan yang tepat dalam berbagai aspek bisnis, mulai dari pengembangan produk baru hingga pengelolaan inventaris. Berikut adalah beberapa manfaat dari *data warehouse*:

1. Analisis bisnis yang lebih unggul

Analisis bisnis yang lebih unggul dapat dicapai melalui *data warehouse* yang memberikan akses kepada para pengambil keputusan untuk mengakses beragam sumber data. Hal ini memungkinkan mereka membuat keputusan berdasarkan informasi yang lebih lengkap dan akurat.

2. Kueri lebih cepat

Data warehouse dirancang khusus untuk menghadirkan kueri dengan kecepatan tinggi, memungkinkan eksekusi data besar dilakukan dengan cepat, baik dengan bantuan tim IT maupun tanpa keterlibatan mereka.

3. Peningkatan kualitas data

Data warehouse meningkatkan kualitas data dengan melakukan proses pembersihan yang dikelola oleh sistem sebelum data dimuat ke dalamnya. Data kemudian diubah menjadi format yang konsisten, mendukung analisis dan pengambilan keputusan berdasarkan data yang berkualitas tinggi dan akurat.

4. Wawasan historis

Adanya data historis dalam *data warehouse* memungkinkan para pengambil keputusan untuk mempelajari tren dan tantangan masa lalu, membuat prediksi, dan mendorong peningkatan bisnis yang berkelanjutan.

5. Data yang lebih aman

Data yang disimpan dalam *data warehouse* lebih aman karena pengaturan sistem keamanan *multi-level* lebih mudah dilakukan dalam penyimpanan data yang terpusat. Selain itu, *data warehouse* juga memungkinkan pemberian

akses terbatas ke data berdasarkan peran dan tanggung jawab karyawan untuk meningkatkan keamanan data perusahaan.

2.2.2. *Nine-Step Methodology*

Berikut ini adalah metodologi perancangan *data warehouse* oleh Kimball yang memiliki sembilan langkah-langkah yang harus dilakukan dalam merancang suatu *data warehouse*:

1. Tahap Pemilihan Proses (*Choosing the Process*)

Dalam perancangan *data warehouse*, tahap ini mengharuskan pemenuhan persyaratan yang matang, seperti analisis pembelian dari berbagai sumber atau dimensi yang berbeda. Oleh karena itu, proses harus dipilih dengan cermat sesuai dengan kebutuhan analisis yang diinginkan.

2. Tahap Pemilihan Sumber (*Choosing the Grain*)

Desain model data harus disesuaikan dengan tabel dan metrik fakta yang ada. Dalam memilih sumber pengetahuan, perlu dipertimbangkan kemampuan mengungkapkan fakta tentang berbagai sumber daya di OLTP (*Online Transaction Processing*). Sebagai contoh, tabel `sales_order_fact` menggambarkan detail pelanggan secara rinci. Pemilihan sumber ini penting karena akan mempengaruhi struktur *data warehouse* dan membantu dalam pengambilan keputusan.

3. Tahap Mengidentifikasi dan Penyesuaian Dimensi (*Identifying and Conforming the Dimensions*)

Dimensi mencerminkan gambaran fakta dalam tabel fakta. Sebagai contoh, setiap catatan dalam tabel `client_dim` mencakup atribut seperti `cust_id`, `cust_number`, `cust_name`, `cust_street_address`, `cust_zip_code`, dan lainnya.

4. Tahap Pemilihan Fakta (*Choosing the Fact*)

Proses ini melibatkan pemilihan sumber data dari tabel fakta untuk menentukan fakta yang akan digunakan. Sebagai contoh, tabel `sales_order_fact` berisi `customer_id`, `order_id`, `product_id`, `order_date`, dan masing-masing menjelaskan fakta dari tabel dimensi yang berbeda, seperti `customer`, `order`, `product`, dan `time`, yang serupa dengan `order_date`.

5. Tahap Menyimpan Perhitungan Awal dalam Tabel Fakta (*Storing Pre-Calculation in The Fact Table*)

Tahap ini terjadi ketika perhitungan tersebut disimpan dalam tabel fakta berdasarkan fakta yang diinginkan.

6. Tahap Melengkapi Tabel Dimensi (*Rounding Out the Dimension Tables*)

Tahap ini bertujuan untuk menambahkan informasi yang lebih rinci ke dalam tabel dimensi agar lebih mudah dipahami oleh pengguna. Level ini akan digunakan untuk analisis oleh pengguna di masa mendatang.

7. Tahap Memilih Durasi Database (*Choosing the Duration of Database*)

Proses ini terkait dengan penyimpanan data dalam *data warehouse*. Sebagai contoh, data yang ingin diambil untuk analisis mencakup periode selama 5 tahun.

8. Tahap Melacak Perubahan pada Dimensi secara Perlahan (*Tracking Slowly Changing Dimension*)

Tahap ini melibatkan tiga jenis perubahan dimensi:

- a. Tipe 1 mencakup perubahan data yang terjadi langsung atau mengupdate tabel dimensi
- b. Tipe 2 mencakup perubahan data yang membentuk record baru dengan surrogate key yang berbeda; dan
- c. Tipe 3 mencakup perubahan data yang akan membentuk atribut atau kolom baru pada tabel dimensi.

9. Tahap Memutuskan Prioritas dan Tipe Query (*Deciding the Query Priorities and The Query Models*)

Tahap akhir ini terkait dengan desain fisik. Masalah utama dalam desain fisik yang mempengaruhi persepsi pengguna akhir tentang data mart adalah urutan tabel fakta yang disusun pada disk, serta agregasi dan ringkasan pencapaiannya.

2.2.3. Basis Data (Database)

Basis data (*database*) adalah suatu sistem yang dirancang untuk menyederhanakan pengaturan, penyimpanan, dan pengambilan data. Basis data terdiri dari kumpulan data yang terorganisir dalam bentuk digital untuk satu atau lebih kegunaan. Pengelolaan basis data dilakukan secara digital melalui sistem manajemen basis data (DBMS), yang bertugas menyimpan konten basis data dan menyediakan fungsi untuk pembuatan data, pemeliharaan data, pengambilan data, dan akses lainnya. Beberapa contoh basis data yang umum digunakan adalah

MySQL, SQL Server, Microsoft Access, Oracle, dan PostgreSQL (Ramadhan and Mukhaiyar 2020).

Adapun beberapa fungsi dari *database* adalah sebagai berikut:

1. Memudahkan identifikasi data dengan mengelompokkannya, seperti menggunakan beberapa tabel atau field yang berbeda.
2. Mengurangi duplikasi data.
3. Meningkatkan pengalaman pengguna dalam berbagai aspek, seperti saat memasukkan data baru.
4. Memori digital dapat menjadi alternatif untuk mengatasi masalah penghematan ruang dalam aplikasi..

2.2.3.1. MySQL

MySQL adalah sebuah implementasi dari sistem manajemen basisdata relasional (RDBMS) yang didistribusikan secara gratis dibawah lisensi GPL (*General Public License*). Setiap pengguna dapat secara bebas menggunakan MySQL, namun dengan batasan perangkat lunak tersebut tidak boleh dijadikan produk turunan yang bersifat komersial. MySQL sebenarnya merupakan turunan salah satu konsep utama dalam basisdata yang telah ada sebelumnya; SQL (Structured Query Language). SQL adalah sebuah konsep pengoperasian basisdata, terutama untuk pemilihan atau seleksi dan pemasukan data, yang memungkinkan pengoperasian data dikerjakan dengan mudah secara otomatis.

MySQL merupakan software *database* open-source yang paling populer di dunia. MySQL menjadi pilihan utama bagi banyak pengembang software dan

aplikasi hal ini dikarenakan kelebihan MySQL diantaranya sintaksnya yang mudah dipahami, didukung program-program umum seperti C, C++, Java, PHP, Python. Pengguna MySQL tidak hanya sebatas pengguna perseorangan maupun perusahaan kecil, namun perusahaan seperti Yahoo!, Google, Nokia, Youtube, Wordpress juga menggunakan DBMS MySQL (Warman and Ramdaniansyah 2018).

Kehandalan suatu sistem basisdata (DBMS) dapat diketahui dari cara kerja pengoptimasi-nya dalam melakukan proses perintah-perintah SQL yang dibuat oleh pengguna maupun program-program aplikasi yang memanfaatkannya. Sebagai peladen basis data, MySQL mendukung operasi basisdata transaksional maupun operasi basisdata nontransaksional. Pada modus operasi nontransaksional, MySQL dapat dikatakan unggul dalam hal unjuk kerja dibandingkan perangkat lunak peladen basisdata kompetitor lainnya (Februariyanti, & Zuliarso, 2012)

2.2.4. *Business Intelligence (BI)*

Business Intelligence adalah rangkaian model matematika dan metodologi analisis yang memanfaatkan data yang ada untuk menghasilkan informasi dan pengetahuan yang berharga untuk mendukung proses pengambilan keputusan yang kompleks. *Business Intelligence* meliputi sebuah arsitektur dan kumpulan operasional yang terintegrasi, serta berfungsi sebagai aplikasi pendukung pengambilan keputusan dan Database yang menyediakan akses mudah bagi komunikasi bisnis terhadap data bisnis (Steven, K., Hariyanto, S., dan Arijanto 2021).

Adapun beberapa tahapan yang ada pada *business intelligence* (Fathurrahman, Yupi Kuspani Putra 2020):

- a. Mengidentifikasi masalah yang memerlukan solusi melalui penggunaan *data warehouse* dan menentukan data yang dibutuhkan untuk menyelesaikan masalah tersebut.
- b. Menentukan sumber data yang diperlukan dan mengambil data tersebut dari sumbernya.
- c. Mengolah data yang diperoleh dari berbagai sumber sehingga menjadi data yang seragam dan konsisten.
- d. Menyimpan data yang telah diolah ke dalam lokasi yang terpusat.
- e. Membangun *data warehouse* dari data yang telah terpusatkan di lokasi tersebut.
- f. Memanfaatkan aplikasi yang memberikan akses ke data dalam *data warehouse* tersebut..

Beberapa elemen bekerja secara bersinergi untuk menciptakan sistem *Business Intelligence*, di mana data organisasi yang ada di *database* operasional, seperti data penjualan, inventaris, dan pelanggan yang diperoleh dari berbagai sistem dan mungkin tidak terhubung satu sama lain, diambil dan dipindahkan ke *Data Warehouse* melalui proses *Extract, Transform, Load* (ETL) (Turban & Volonino, 2011). Kemudian dalam *Data Warehouse*, data-data tersebut dihubungkan satu dengan yang lainnya dan membentuk *database* multidemensi yang kemudian akan diolah menjadi informasi yang berguna, baik itu berupa Graph,

Report, Alert, Forecasting, Data Mining, atau Query sesuai dengan kebutuhan organisasi.

2.2.4.1. *Microsoft PowerBI*

Power BI adalah sebuah aplikasi *Business Intelligence* yang dimiliki oleh Microsoft yang memiliki kemampuan untuk menampilkan data dalam bentuk visualisasi, memungkinkan pembuatan query, koneksi data, dan laporan (Ronald, 2008). Keunggulan Power BI terletak pada antarmuka pengguna yang mudah digunakan, mirip dengan aplikasi Microsoft Office, serta kemampuannya untuk mengolah data dengan detail dan menyajikannya dalam bentuk yang lebih interaktif. Power BI dapat diakses melalui tiga platform, yaitu platform desktop yang bisa diinstall di komputer atau laptop, platform web yang dapat diakses melalui web browser, dan platform mobile yang bisa digunakan di smartphone.

Dengan bantuan Power BI, pengguna dapat menganalisis data dan mengambil keputusan bisnis secara lebih efisien di perusahaan, karena data yang relevan dapat dengan mudah diakses dan digunakan untuk memecahkan masalah. Power BI menyediakan berbagai kemampuan analisis, termasuk Predictive analytics, Data Visualizations, R integration, dan Data Analysis Expression. Dalam aplikasi Power BI, terdapat tiga konsep sistem kerja, yaitu Datasets, Report, dan Dashboard (Mahendra, Sukawirasa, Dimastawan Saputra, & Mahendra, 2021)

2.2.4.2. Dashboard

Dashboard adalah representasi visual dari informasi yang sangat relevan dan penting untuk mencapai satu atau lebih tujuan, yang dapat ditampilkan dalam satu layar sehingga memudahkan pengguna dalam memantaunya. Perguruan tinggi memerlukan berbagai jenis dashboard untuk mendukung upaya dalam penjaminan mutu dan peningkatan kualitas (Januarita and Dirgahayu 2015).

2.2.5. Online Analytical Processing (OLAP)

OLAP (*On-Line Analytical Processing*) adalah suatu metode khusus yang digunakan untuk melakukan analisis data yang terdapat pada media penyimpanan data (*database*) dan kemudian membuat laporannya sesuai dengan permintaan user. Istilah OLAP muncul pertama kali pada tahun 1993 yang diperkenalkan oleh Edgar F. Codd, S. B. Codd, dan C. T. Salley dalam dokumen untuk Arbor berjudul “*Providing OLAP to User-Analysts: An IT Mandate*”. OLAP menggambarkan sebuah klas teknologi yang dirancang untuk analisa dan akses data secara khusus. Apabila pada proses transaksi pada umumnya semata-mata adalah pada relational *database*, OLAP muncul dengan sebuah cara pandang multidimensi data. Cara pandang multidimensi ini didukung oleh teknologi multidimensi database. Cara ini memberikan tehnik dasar untuk kalkulasi dan analisa oleh sebuah aplikasi bisnis (Hendro Poerbo Prasetiya, 2016).

2.2.6. *Extract, Transform and Load (ETL)*

ETL merupakan sebuah proses-proses yang ada didalam sebuah *data warehouse*. Proses ETL yaitu: mengekstrak (extract) data yang akan digunakan dari sumber eksternal (source), merubah (transform) data yang akan digunakan sesuai dengan keperluan bisnis, memasukkan (load) data yang akan digunakan ke *data warehouse* (Hasan and Febriandirza 2021). Proses dari *Extract, Transform and Load (ETL)* merupakan sebuah proses integrasi data yang mengekstrak data dari sumber luar, mengubah data sesuai dengan kebutuhan bisnis dan menyimpannya kedalam *data warehouse*. Data yang digunakan didalam proses ETL pun bisa didapatkan dari berbagai jenis sumber termasuk dari sistem aplikasi ERP (Enterprise Resources Planning), flat file model, dan spreadsheet (Santoso and Yulia 2017).

2.2.6.1. *Pentaho Data Integration Kettle*

Pentaho Data Integration (PDI) atau Kettle merupakan perangkat lunak dari Pentaho yang dapat digunakan untuk proses ETL (*Extraction, Transformation, dan Loading*). PDI berguna untuk migrasi data, membersihkan data, dan melakukan loading dari file ke dalam database, atau sebaliknya, dalam volume besar. PDI menyediakan antarmuka pengguna grafis dan komponen *drag-and-drop* yang mempermudah pengguna. Elemen utama dari PDI terdiri dari *Transformation* dan *Job*. *Transformation* merupakan kumpulan instruksi yang digunakan untuk mengubah input menjadi output yang diinginkan (input-proses-output). Sementara itu, *Job* merupakan kumpulan instruksi yang menjalankan transformasi tersebut.

PDI memiliki tiga komponen utama, yaitu *Spoon*, *Pan*, dan *Kitchen*. *Spoon* berfungsi sebagai antarmuka pengguna untuk membuat *Job* dan *Transformation*. *Pan* adalah alat yang berfungsi untuk membaca, mengubah, dan menulis data. Sedangkan *Kitchen* adalah program yang digunakan untuk mengeksekusi job.