|  |  |  |
| --- | --- | --- |
| C:\Users\Admin\Pictures\JATA KPM.png  **BAHAGIAN PENDIDIKAN DAN LATIHAN TEKNIK VOKASIONAL**  **KEMENTERIAN PENDIDIKAN MALAYSIA**  **ARAS 5 & 6, BLOK E14, KOMPLEKS E,**  **PUSAT PENTADBIRAN KERAJAAN PERSEKUTUAN**  **KERTAS PENERANGAN**  ***(INFORMATION SHEET)*** | | |
| **KOD DAN NAMA PROGRAM NOSS** | IT-010-3:2016 PEMBANGUNAN APLIKASI | |
| **TAHAP NOSS** | 3 | |
| **KOD, NAMA CU DAN WA NOSS** | **CU04/WA2 - PERFORM DATA POPULATION TO DEVELOPMENT SERVER** | |
| **KOD DAN NAMA PROGRAM KV** | SISTEM PENGURUSAN PANGKALAN DATA DAN APLIKASI WEB | |
| **KOD DAN NAMA KURSUS KSKV** | K  KPD 3033 DEVELOPMENT ENVIRONMENT DEPLOYMENT | |
| **NO.DAN TAJUK KOMPETENSI** | K  K2 PERFORM DATA POPULATION TO DEVELOPMENT SERVER | |
| **NO. KOD NOSS** | IT-010-3:2016 - C04/P(6/12) | Muka Surat : 1 Drp : 7 |
| **NO. KOD KSKV** | KPD 3033 / P(6/12) |

**TAJUK:**

**JENIS DATA MIGRASI**

**TUJUAN:**

Kertas penerangan ini adalah bertujuan menerangkan mengenai :

1. Terangkan jenis data migrasi konsep *data extraction*.
2. Terangkan jenis data migrasi konsep *data loading*.
3. Terangkan jenis data migrasi konsep *data massage/ data cleaning*.

**PENERANGAN/***INFORMATION* **:**

1. **JENIS DATA MIGRASI**

Definisi data migrasi.

* Data migrasi ialah **proses pemindahan data antara komputer, peranti penyimpanan atau format**. Ia adalah kunci utama untuk pelaksanaan, peningkatan (*upgrade*) atau untuk pelaksanaan. Pemindahan yang berlaku adalah melalui perisian program atau skrip yang digunakan untuk pemindahan data secara automatik.
* Data migrasi dikategorikan sebagai *storage migration*, *database migration*, *application migration* dan migrasi proses perniagaan.

Secara umum migrasi data dapat memudahkan pengguna dalam menyelesaikan masalah berikut :

1. Menggabungkan sistem dari kedua sistem yang berbeza menjadi satu sistem baru.
2. Bermigrasi ke salah satu sistem yang lain.
3. Menggunakan sistem sedia ada kemudian membangunkan sebuah  *datawarehouse* diatasnya.

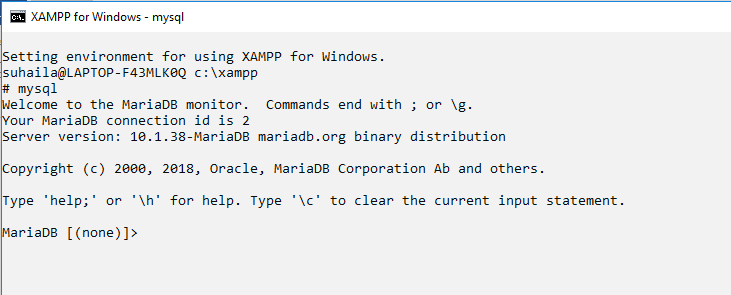
Terdapat beberapa kaedah yang boleh digunakan untuk melaksanakan migrasi data:

1. Export dan Import

Dalam beberapa kes, terdapat alat yang menyokong eksport data ke pangkalan data lain dan kemudian boleh di import dalam format yang sama. Sebagai contoh, perisian perniagaan membolehkan data untuk diimport daripada fail teks terhad. Ini memudahkan para penganalisis data untuk bekerja dengan data dalam hamparan, untuk mendapatkannya dalam format yang diharapkan dan kemudian dapat mengimport data.

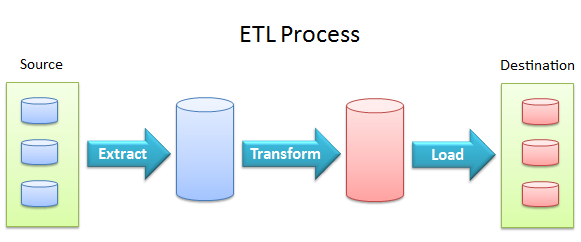
b. Script

Skrip pangkalan data adalah cara yang biasa digunakan untuk mengimport dan mengeksport data dari pangkalan data. Pada asasnya, kaedah ini adalah sama dengan proses import / eksport dengan menggunakan antara muka pengguna. Jika data yang digunakan cukup kecil, hamparan boleh digunakan untuk menyediakan data untuk pangkalan data sasaran. Sebagai alternatif, operasi pangkalan data seperti arahan SQL sering digunakan untuk membuat jadual sementara dan memindahkan data sekeliling agar sesuai dengan model data baru.



Rajah 1: script adalah menggunakan Mysql Command Line

c. Extract, Transform, Load (ETL)



Rajah 2: Ilustrasi proses ETL

Merupakan **proses atau ciri yang disediakan oleh beberapa alat untuk memindahkan data**. Alat ETL ini biasanya boleh **mengendalikan set data besar**. Alat ini direka untuk menyambung kepada beberapa sumber data dan secara automatik akan melaksanakan transformasi seperti pemetaan dan peraturan perniagaan untuk menukar dari satu model data kepada yang lain**.**

**KONSEP EXTRACT, TRANSFORM, LOAD (ETL)**

1. Data extraction

* Data extraction ialah **satu proses analisis data yang diperolehi dari pelbagai sumber.**
* Pemprosesan data yang dilakukan akan melibatkan penambahan metadata dan integrasi dengan data lain, atau proses lain dalam proses kerja.
* Pengekstraktan (*ectraction*) data biasanya diperolehi daripada data yang tidak berstruktur dan format yang berbeza.
* Terdapat pelbagai bentuk format yang tidak berstruktur seperti jadual, indeks dan analitik.
* Proses pengekstraktan data melibatkan pengambilan semula data dari sumber data lain dan akan diekstrak (diubah) ke dalam bentuk yang dikehendaki. Seterusnya akan di muat naik ke dalam pangkalan data.

1. Data loading

* Data loading digunakan dalam teknik pengekstraktan (extraction) dan pemuatan (loading) berdasarkan pangkalan data (databases).
* **Data akan di muatkan ke dalam satu aplikasi dengan format yang berbeza daripada sebelumnya**.
* Pemuatan data ialah **proses penyalinan** dan **pemuatan data** atau set data dari satu fail sumber, folder atau aplikasi ke pangkalan data atau aplikasi serupa.
* Ia biasanya dilaksanakan dengan menyalin data digital dari sumber dan **menyisipkan atau memuatkan data** ke penyimpanan data atau utiliti pemprosesan.
* Contoh, data yang disalin daripada fail ke aplikasi pangkalan data, dan format akan diubah daripada .doc atau .txt kepada format .CSV atau DAT.
* Data loading biasa dijalankan diperingkat akhir proses ETL.

1. Data massage / data cleaning

* Semasa proses transformasi data, beberapa prosedur perlu dilaksanakan ke atas data yang telah di *extract* supaya data yang di hasilkan di akhir proses pemuatan (*loading)* adalah mengikut target yang diharapkan.
* Proses ini hanya melaksanakan “pembersihan data” kerana hanya data yang diperlukan sahaja yang akan dimuat naik.
* Isu utama semasa pembersihan:
  + data yang telah bersih adalah mengikut spesifikasi sistem dari segi data type, set karakter.

contoh:

* + Jantina : label Perempuan kepada P
* Proses menukar data atau maklumat daripada satu format ke format lain, biasanya dari format sistem sumber ke format yang diperlukan bagi sistem yang baru.

Aktiviti dalam ETL merangkumi:

1. pemulaan kitaran ETL

2. Membina data rujukan

3. E*xtract* dari sumber

4. Pengesahan data yang telah di *extract* dari sumber

5. *Tranform* – merangkumi proses pembersihan, rujuk peraturan syarikat, semak integrity data, membina agreget atau membuat agreget data)

6. Memuat naik dalam *Staging* (jika prosedur ini diperlukan)

7. Membina laporan – menyatakan peraturan yang digunakan, laporan kerosakan –jika ada)

8. Hantar – memuat naik dalam sistem yang disasarkan.

9. Arkib – simpan fail tersebut dalam storan utama (*backup*)

d. Integrasi

Jika integrasi data dibangunkan antara dua sistem, ia digunakan untuk memindahkan data awal. Alat integrasi seperti ESB (Bus Service Enterprise) mempunyai keupayaan pemetaan dan transformasi yang boleh mengendalikan penghijrahan data.

Alat yang direka untuk integrasi masa nyata biasanya dianggap agak intensif untuk penghijrahan data satu kali. Walau bagaimanapun, mereka berpotensi untuk melakukan pekerjaan itu.

e. Migrasi pangkalan data   
Proses ini hanya melibatkan permindahan data dari satu pangkalan data ke pangkalan data yang lain. Walaubagaimanapun, terdapat isu yang mungkin timbul contoh:

* Data type yang tidak sama atau salah
* Set karakter yang berbeza

f. Migrasi aplikasi

Proses ini akan melibatkan perpindahan data antara aplikasi ke aplikasi yang lain. Hal ini mungkin boleh menjadi lebih komplek jika sesuatu aplikasi tidak di sokong oleh aplikasi yang ingin digunakan untuk proses migrasi.

Contoh situasi migrasi aplikasi:

i. Membina data menggunakan Microsoft Excel dan menyimpan dengan   
format .CSV

ii. Kemudian memuat naik ke dalam PhpMyAdmin.

Oleh itu, dapat disimpulkan, dengan menggunakan proses migrasi data ini, apabila kita memindahkan sistem lama ke sistem baru dari suatu aplikasi, kita boleh menggunakan data sedia ada dari sistem terdahulu untuk sistem baru dengan beberapa kaedah atau kaedah proses sebahagiannya disokong oleh alat yang ada. Kelebihan dari proses pemindahan data ini adalah bahawa ia dapat menjalankan proses pemindahan program secara programatik, kecekapan waktu, dan mengurangi kesalahan manusia.

**SOALAN /** *QUESTIONS :*

1. Berikan 3 jenis data migrasi
2. Terangkan konsep Extract, Transform, Load (ETL)

**RUJUKAN /** *REFERENCES :*

1. Keith Cooper, Linda Torczon(2011), Engineering a Compiler (2nd Edition), Morgan Kaufmann, ISBN: 978-0120884780
2. Steve McConnell (2004), Code Complete: A Practical Handbook of Software Construction, (2nd Edition),Microsoft Press, ISBN: 978-0-7356-1967-8
3. Michael L. Scot (2009) , Programming Language Pragmatics, Third Edition, Morgan Kaufmann, ISBN-13: 978-0123745149
4. Simon Marlow (2013), Parallel and Concurrent Programming in Haskell: Techniques for Multicore and Multithreaded Programming (1st Edition), O'Reilly Media, ISBN-13: 978-1449335946