

6. Implementation

6.1. Introduction

After finalizing the design, the implementation needs to be done according to the process model which was selected.

6.2. Physical database design

The physical database design of the MIS is made up of three sets areas belonging to three stages. They can be described as the loading area, Staging area and the Production (physical) area. Even though the all the system are supported by Oracle databases including the MIS, the data from the source OLTP systems are written to text files and then loaded to the Loading area of the MIS. The reason for doing so, instead transferring data between tables is to avoid invalidating the mappings, which would happen if the Source system data structures change due to a version change. We have avoided this RISK by using this mechanism where if a structural change occurs in the source system it becomes the responsibility of the vendor to still provide the data in required text file format. We easily get into this agreement since the vendor does not want to divulge information regarding internals of his package to us (the client).

Next the data is loaded to the staging area where it is transformed into clean data using many routines. Value addition happens to the data in form of categorization.

Transformed data is then loaded into database.

The main concern was the process of cleaning the data.

- DFCC bank started its business in 1956, at that point of time there were no systems thus the recorded data were not in a very consistent form.
- The National Identity Card (NIC) was introduced somewhere in early 1970's
- So the earlier data did not have the NIC or a Company registration number as reference. Whereas now it is mandatory to put the NIC or the Passport number for an individual and the company registration number for a company.

- For the data which got discarded were scrutinized and the bank officers were instructed to find any available information so that they can be included correctly on to the system.
- When typing names the consistency was not maintained. Some of the users have typed as surname first whereas some had the surname coming in the latter part of the name.
- For the location dimension the city was derived from the client details. This also had issues as people spell the names of cities differently. Eg 'Matara' or 'Mathara'. In order to rectify this, the table provided by the Postal department for the purpose of postal codes was used.

6.3. Data cleansing

The typing errors in the client information table were the hardest to clean. The `global_id` is the column where the NIC or the company registration number is saved. Out of the 39000 odd clients/companies there was 8470 number of records that didn't have it saved. The following table (table 6.1) tabulates some of the scenarios generated for the `global_id` column to identify the discrepancies. The relevant SQL statements are in Appendix B. As the `global_ID` field contains the NIC/Passport NO/Company reg. no it is the main column from which the data from the 3 systems are to be merged this column needs to clean in order to get an accurate picture of a client.

Scenario 1:

GLOBAL_ID	COUNT(*)
global id is NIC and length is 11 and not like 9 no_space_V or X	157
global_id has spaces in between	1273
global_id has spaces in front	72
global_id is not a valid nic, length too long	2037
global_id is null	8470
total number of clients	39504

Table 6-1: Customers with erroneous Global ID 1

To elaborate more the following query gave an output that showed the errors that a data entry person makes.

Scenario 2:

```
select client_no, global_id, Client_short from dfcc.fm_client
where global_id like '%XX%'
```

CLIENT_NO	GLOBAL_ID	CLIENT_SHORT
000000049552	XXXXXXXXXX	LINK GROUP
000000018608	XXXXX	WAYAMBA DEVELOPMENT BANK
000000045507	XXXX	XXXCHL SALGADO
000000104951	XXXXXXXXXX	MAHESHI
000000107650	XX02/1783	R.M.K.ENGINEERING
000000106042	XXXHG/E8/BU/1643	XXXXXXJANANI ENTERPRISES
0000IH003257	XXXX	H.M.D.S.K. HERATH
000000106742	XXXXX	SEEKSHANA MOTORS GROUP
000000013782	XCXCXX	THE SWADESHI HE

Table 6-2: Customers with erroneous Global ID II



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Scenario 3:

Following query displays the errors found in the city field. This is also a required field as from this column a dimension has been created. It was observed that there were many inconsistencies in the entered data. These were also updated with the help of the executives of the specified branch.

```
select client_city
from dfcc.fm_client c
where client_city is not null and
upper( c. client_city) not in (select upper(city) from dfcc.postal_codes )
order by client_city
```

The output – extract

```
CLIENT_CITY
1 ANURADHAPURA
1 ANURADHAPURA
Ampitiya
Aniwatta kandy
```

Table 6-3: Customers with erroneous city details

In-order to rectify the errors the executive who has entered the data was identified and instructed to update the details.

6.4. Administration interface

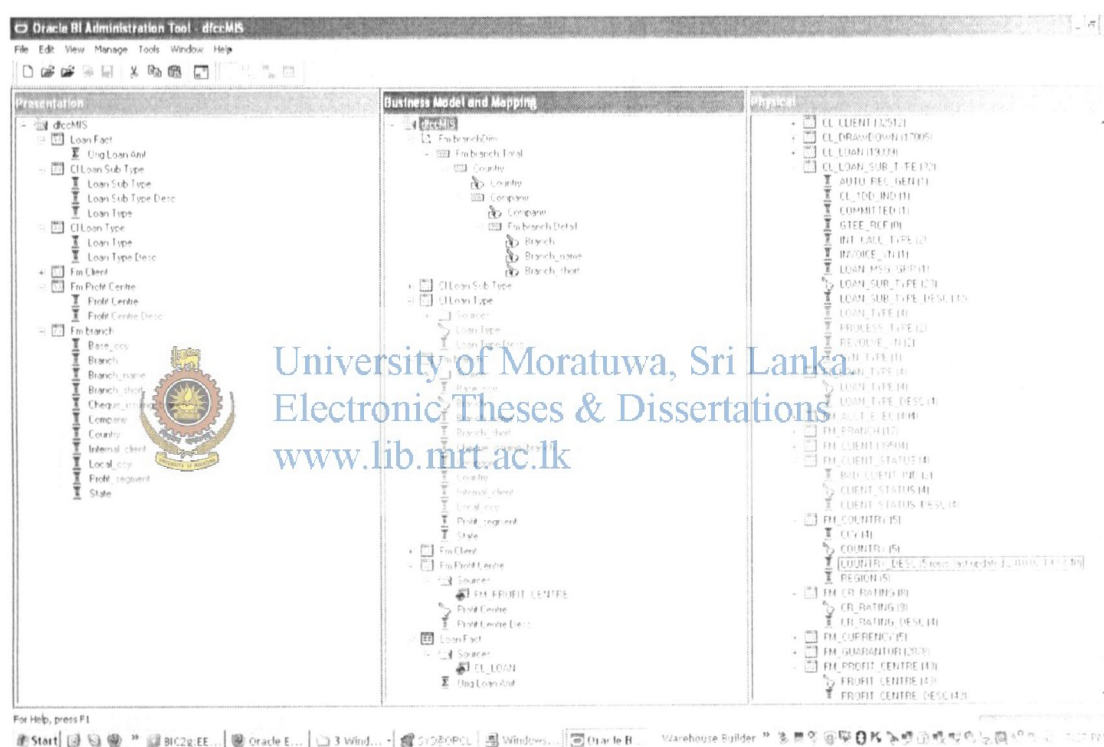


Figure 6-1: Oracle BI Administration tool

From the cleansed data area the Physical layer imports the data. It is then mapped to the Dimensional hierarchies with labels renamed so that could be understandable and then pushed on to the presentation layer where the users can access the data via Oracle BI Answers or Dashboards. BI administrator tool is shown in figure 6.1.

6.5. Security interface

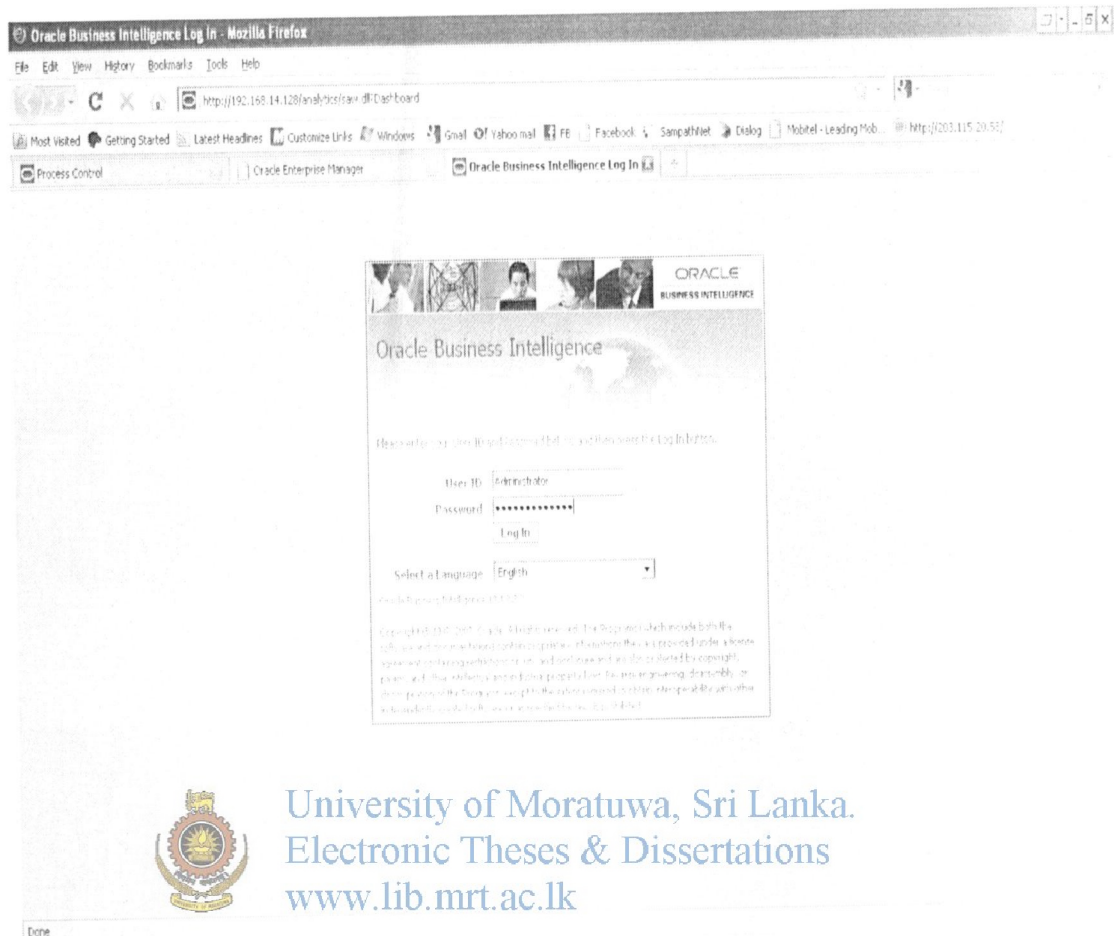


Figure 6-2: User authentication

User creation and authentication needs to be done through BI home page.

6.6. User interface

After the login screen (figure 6.2) the user is taken to the home page (figure 6.3).

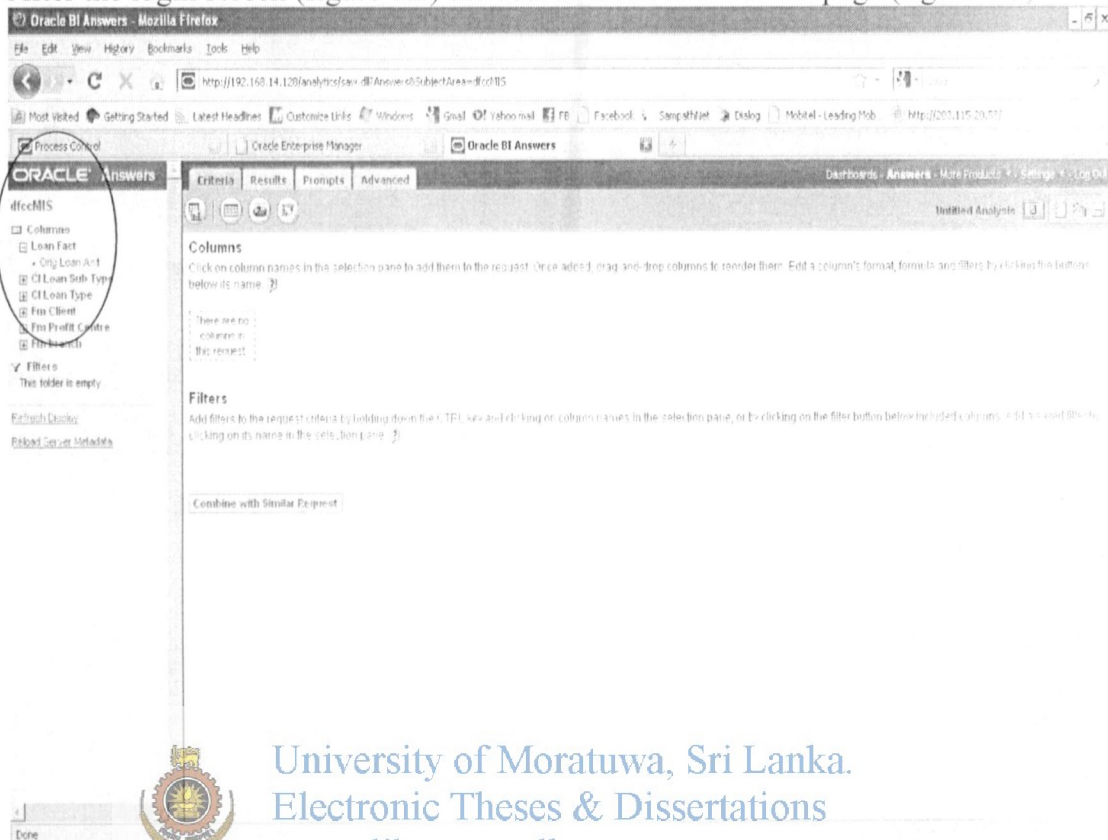


Figure 6-3: User home page

The blue oval in the figure 6-3 shows the columns which are available for the user to generate reports. Sample reports are in Appendix C.