BI

Mini Project

Census Reporting System

IGATE
Speed.Agility.Imagination

Document Control

Reference No:			
Security Classification:	iGATE Internal		
Issue Date:			
Author(s):	**Name**		**Title**
	Rajita Dhumal		Dy Manager
Reviewer(s):	**Name**		**Title**
Issuer(s):	**Name**		**Title**
	Rajita Dhumal		Dy Manager
Issuer Signature(s):			
Distribution:	**Name**		**Title**
	iLEARN		All Designations in iLEARN
Document History:	**Date**	**Revision**	**Change**
	Apr 2014	1.0	Document creation

Table of Contents

1 INTRODUCTION

This document outlines a mini project for the BI LOT. The project is to develop Census Reporting System.
This document contains the work flow of the system and gives guidelines on how to build the functionality gradually in each of the course modules of the BI LOT.

1.1 SETUP CHECKLIST FOR MINI PROJECT

Minimum System Requirements

- Intel Pentium 90 or higher (P166 recommended)
- Microsoft Windows 95, 98, or NT 4.0, 2k, XP, Windows 2007.
- Memory: 32MB of RAM (64MB or more recommended)
- Internet Explorer 6.0 or higher
- Oracle 9i client and access to oracle 9i server
- Informatica 9.5.1
- BOXI R4

1.2 INSTRUCTIONS

- The code modules in the mini project should follow all the coding standards.
- Create a directory by your name in drive **<drive>**. In this directory, create a subdirectory **MiniProject**. Store your Project here.
- You can refer to your course material.
- You may also look up the help provided in the BI docs and documentation provided with Informatica and Business Objects.

2 PROBLEM STATEMENT

2.1 OBJECTIVE

Development of a **Census Reporting System.**

2.2 ABSTRACT OF THE PROJECT

A Census is the procedure of systematically acquiring and recording information about the members of a given population. It is a regularly occurring and official count of particular population. The Population Census is the most detailed information source on the population at the level of small localities and small groups in population. The data acquired forms the basis of information that is available to public and private elements at the national and local level for the purpose of decision making in variety of areas of the life of residents of the cities.

The purpose of this project is to gather information about general population in order to present a full and reliable picture of population in Cities- its Demographic and housing conditions, social and economic characteristics. This project is essential for City wise comparisons of any kind of Statistics. It collects data on many attributes of population, not just for the count of people, although population estimates remain an important function of Census

2.3 TECHNOLOGY USED:

- Oracle 9i (Database)
- Informatica power center 9.5.1 (ETL Tool):-
- BusinessObjects XI Release 4

3 IMPLEMENTATION IN BI LOT

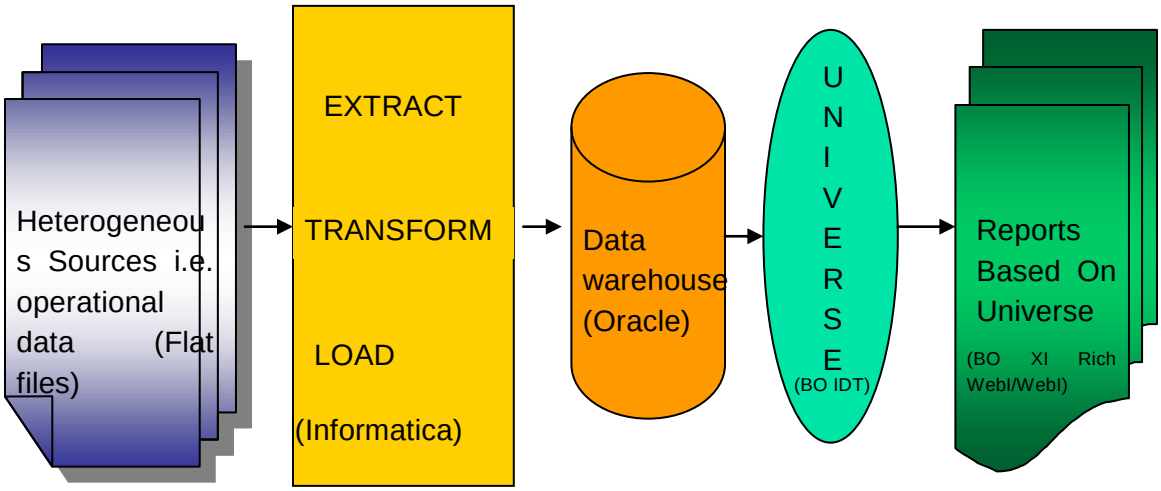
3.1 SUMMARY OF THE FUNCTIONALITY TO BE BUILT:

The participants need to develop the **Census Reporting System** by building the functionality incrementally in each of the course modules of BI LOT.

Sr. No	Course	Duration (in PDs)	No. of Saturdays	Functionality to be built
1.	Programming Foundation using Pseudo code	5	2	
3.	Oracle	7		
4.	Unix	4		
5.	Oracle Test	1		
6.	Informatica 9.5.1	6	1	Perform ETL to generate the fact tables and Dimension tables.
6.	Informatica 9.5.1 Test	1		
7	BOXI	5	1	Create a Universe in BO and create Reports
8	BOXI Test	1		
9	OBIEE 11g	5	1	Create a package in framework Manager and create reports in Report Studio and Analysis Studio
10	OBIEE 11g Test	1		

3.2 GUIDELINES ON THE FUNCTIONALITY TO BE BUILT:

Project flow



Schema Design:

The project follows the start schema approach.

- The Dimension tables are :
  - City
  - Education
  - Occupation
- Fact Tables are :
  - Fact_Census Occupation
  - Fact_Census Education

FACT_CENSUS_OCC	
Person_id	Varchar
Door_num	Number(p,s)
Gender	Varchar
DOB	Varchar
Maritial_status	Varchar
Religion	Varchar
Caste	Varchar
Disability	Varchar
Type_of_disability	Varchar
status_of_attendance	Varchar
Occupation_id	Varchar
Occupation	Varchar
Nature_of_industry	Varchar
sector	Varchar
Salary	Number(p,s)
Children_surviving	Varchar
Children_ever_born	Varchar
Persons_expired_in_10yr	Varchar
Type_of_house	Varchar
PAN_card	Varchar
Ration_card	Varchar
Adhar_card	Varchar
Area	Varchar
City_id	Varchar
Literacy_status	Varchar
Mode_of_transport	Varchar
City_name	Varchar

DIM_EDUCATION	
Education_id	Varchar
Highest_education	Varchar

DIM_OCCUPATION	
Occupation_id	Varchar
Occupation_name	Varchar

DIM_CITY	
City_id	Varchar
City_name	Varchar

FACT_CENSUS_EDU	
Person_id	Varchar
Door_num	Number(p,s)
Gender	Varchar
DOB	Date/time
Maritial_status	Varchar
Religion	Varchar
Caste	Varchar
Disability	Varchar
Type_of_disability	Varchar
status_of_attendance	Varchar
Education_id	Varchar
Highest_Education	Varchar
Nature_of_industry	Varchar
sector	Varchar
Salary	Number(p,s)
Children_surviving	Varchar
Children_ever_born	Varchar
Persons_expired_in_10yr	Varchar
Type_of_house	Varchar
PAN_card	Varchar
Ration_card	Varchar
Adhar_card	Varchar
Area	Varchar
City_id	Varchar
Literacy_status	Varchar
Mode_of_transport	Varchar
City_name	Varchar

Note: The datatype/length for the Dimension/Fact table attribues can be changed as required. Additional fields can be added, if required.

- **City:**
  This dimension contains information about the City on which Census Reports are based
  This CSV file comprises of the following Fields:

- City_Id
- City_Name

There are 3 Cities namely Mumbai, Bangalore and Pune.

- **Education**

This dimension contains Highest Education details and the CSV file comprises of the following Fields:

- Education_ID
- Highest_Education

The Highest Education Levels are :

- ✓ SSC(Secondary School Certificate),
- ✓ HSC(Higher Secondary School),
- ✓ UG(Undergraduate),
- ✓ PG(Postgraduate)

The Facts are Fact_Census_Occ, Fact_Census_Edu.

- Fact_Census_Occ

This is a Fact table that contains the batting and bowling related information of each player and the country which he belongs to.

- ✓ Person_id
- ✓ Door_number
- ✓ Gender
- ✓ DOB
- ✓ Maritial_status
- ✓ Religion
- ✓ Caste
- ✓ Disability
- ✓ Type_of_disability
- ✓ Status_of_attendance
- ✓ Occupation_id
- ✓ Occupation
- ✓ Nature_of_industry
- ✓ Sector
- ✓ Salary
- ✓ Children_surviving
- ✓ Children_ever_born

- ✓ Persons_expired_in_10yr
- ✓ Type_of_house
- ✓ PAN_card
- ✓ Ration_card
- ✓ Adhar_card
- ✓ Area
- ✓ City_id
- ✓ Literacy_status
- ✓ Mode_of_transport
- ✓ City_name

■ **Fact_Census_Edu.**

This is a Fact table that would contain the batting and bowling related information of each player and the country which he belongs to.

- ✓ Person_id
- ✓ Door_number
- ✓ Gender
- ✓ DOB
- ✓ Maritial_status
- ✓ Religion
- ✓ Caste
- ✓ Disability
- ✓ Type_of_disability
- ✓ Status_of_attendance
- ✓ Education_id
- ✓ Highest_Education
- ✓ Nature_of_industry
- ✓ Sector
- ✓ Salary
- ✓ Children_surviving
- ✓ Children_ever_born
- ✓ Persons_expired_in_10yr
- ✓ Type_of_house
- ✓ PAN_card
- ✓ Ration_card
- ✓ Adhar_card

- ✓ Area
- ✓ City_id
- ✓ Literacy_status
- ✓ Mode_of_transport
- ✓ City_name

Sample Reports

1. City wise population Density
2. Gender wise population Density
3. Birth Rate -1) Male 2)Female 3) Children
4. Death Rate- 1) Male 2)Female 3) Children
5. Mode of transport to workplace
6. Working Women – 1) Literate 2) Illiterate
7. Working Men – 1) Literate 2) Illiterate
8. Occupation wise classification
9. Education wise classification (SSC,HSC,UG &PG)
10. Religion(Hindu, Muslim, Christian)
11. Job wise classification (government, private & self employed)

Referred to the shared source files.

4 EVALUATION AND ASSESSMENT PARAMETERS

This mini project will be done in groups of five or six. Each group will identify a Team Lead who will decide which team member will code for which functionality. This project shall be evaluated in two phases. Evaluation will be done using online presentation mode, where each group will present their respective modules.

Evaluation Parameters (100 Marks)

1. **ETL Scripts and DW design (25)**
   The extent to which the ETL was performed, the transformations used
   Deliverables in terms of time lines
2. **Universe Design (35)**
   The way the universe was designed with appropriate classes and objects, conditions created
3. **Deliverables(Reports) (40)**
   Presentation of reports.