

# **STUDENT REPORT**

*Project report submitted in partial fulfillment of the requirement for  
the degree of*

## **BACHELOR OF TECHNOLOGY**

**IN**

### **ELECTRONICS AND COMMUNICATION ENGINEERING**

By

**DEEPANKAR MATHUR**

**151064**

UNDER THE GUIDANCE OF

**K N MONISH**



JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

JUNE 2019

## ACKNOWLEDGEMENT

I take this opportunity to express my sincere gratitude to the Director, Jaypee University of Information Technology, Solan for providing this opportunity to carry out the present work.

The constant guidance and consolation got from **Dr Meenakshi Sood** Assistant Professor (Senior Grade) Department of Electronics and Communication Engineering has been of incredible assistance in conveying our present work and helped us in finishing this task with success.

I would like to express a deep sense of gratitude to my Project Advisor **Mr. K N Monish**, Project Manager and **Mohd. Shafaf Moosa**, for their guidance and support in defining the design problem and towards the completion of my project work. Without their wise counsel and able guidance, it would have been impossible to complete the thesis in this manner.

I am also thankful to our batch owners **Ms. Eypsita Pradhan & Ms. K Girija** and Academy heads **Ms. Sumati Vinay Narasimhan** and **Ms. Bhuvana Kumar** for their intellectual support throughout the course of this work.

Finally, I thank my parents and my family members for giving the moral support and abundant blessings in all of activities and my dear friends who helped me to endure the difficult times with their unfailing support and warm wishes.



## JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

(Established by H.P. State Legislative vide Act No. 14 of 2002)  
P.O. Wagnaghat, Teh. Kandaghat, Distt. Solan - 173234 (H.P.) INDIA  
Website: [www.juit.ac.in](http://www.juit.ac.in)  
Phone No. (91) 01792-257999  
Fax: +91-01792-245362

### CERTIFICATE

This is to certify that the work reported in the B.tech project report entitled “**Student Report**” which is being submitted by **Deepankar Mathur** in fulfillment for the award of Bachelor of Technology in Electronics and Communication Engineering by the Jaypee University of Information Technology, is the record of candidate’s own work carried out by him/her under my supervision. This work is original and has not been submitted partially or fully anywhere else for any other degree or diploma.

-----  
**Dr. Meenakshi Sood**

Assistant Professor (Senior Grade)  
Department of Electronics & Communication Engineering  
Jaypee University of Information Technology, Wagnaghat,

**JUIT**  
विद्या तत्व ज्योतिसमः

## **DECLARATION BY THE SCHOLAR**

I hereby declare that the work reported in the B-Tech thesis entitled **STUDENT REPORT** submitted at **Jaypee University of Information Technology, Waknaghat India**, is an authentic record of my work carried out under the supervision of K N Monish I have not submitted this work elsewhere for any other degree or diploma.

Deepankar Mathur

Department of Electronics and Communication

Jaypee University of Information Technology

25<sup>th</sup> May 2019

# TABLE OF CONTENTS

<b>DECLARATION BY THE SCHOLAR</b>	<b>i</b>
<b>ACKNOWLEDGEMENT</b>	<b>ii</b>
<b>LIST OF ACRONYMS AND ABBREVIATION</b>	<b>v</b>
<b>LIST OF FIGURES</b>	<b>vi</b>
<b>LIST OF TABLES</b>	<b>vii</b>
<b>CERTIFICATE</b>	<b>viii</b>
<b>ABSTRACT</b>	<b>ix</b>
<b>CHAPTER 1</b>	<b>1</b>
<b>INTRODUCTION</b>	<b>1</b>
<b>1.1 ETL</b>	<b>1</b>
1.1.1 How ETL Works?	2
1.1.2 Informatica Tool	2
1.1.3 Architecture	2
1.1.4 Client Components in Informatica	3
1.1.5 Sources & Targets	4
1.1.6 Informatica Transformation Overview	5
<b>1.2 Scope</b>	<b>6</b>
<b>1.3 Assumptions, Dependencies &amp; Constraints</b>	<b>7</b>
<b>CHAPTER 2</b>	<b>8</b>
<b>DATA SOURCES AND ACCESS MECHANISMS</b>	<b>8</b>
<b>CHAPTER 3</b>	<b>9</b>
<b>SOURCE SYSTEM CLEANING RULES</b>	<b>9</b>
<b>3.1 Department Dimension</b>	<b>9</b>
<b>3.2 Student Dimension</b>	<b>9</b>
<b>3.3 Marks Fact Dimension</b>	<b>9</b>
<b>3.4 Subject Aggregate Dimension</b>	<b>9</b>
<b>3.5 Subject_Pass_Percentage Dimension</b>	<b>10</b>
<b>CHAPTER 4</b>	<b>11</b>
<b>TARGET SOURCE MAPPING AND TRANSFORMATION RULES</b>	<b>11</b>
<b>4.1 Department Dimensions</b>	<b>11</b>
<b>4.2 Student Dimension</b>	<b>11</b>
<b>4.3 Marks Fact Dimension</b>	<b>12</b>
<b>4.4 Subject Aggregate Dimension</b>	<b>13</b>
<b>4.5 Subject_Pass_Percentage_Aggregate Dimension</b>	<b>14</b>
<b>CHAPTER 5</b>	<b>15</b>
<b>ETL DESIGN OVERVIEW AND COMPONENTS</b>	<b>15</b>
<b>CHAPTER 6</b>	<b>17</b>

<b>ETL PROGRAM SPECIFICATIONS</b>	<b>17</b>
<b>6.1 Customer Dimension</b>	<b>17</b>
6.1.1 Source	17
6.1.2 Target	17
6.1.3 Data Flow Diagram	17
<b>6.2 Student Dimension</b>	<b>18</b>
6.2.1 Source	18
6.2.2 Target	18
6.2.3 Data Flow Diagram	18
<b>6.3 Marks_Fact</b>	<b>19</b>
6.3.1 Source	19
6.3.2 Target	19
6.3.3 Data Flow Diagram	20
<b>6.4 Subject_Average_Aggregate</b>	<b>21</b>
6.4.1 Source	21
6.4.2 Target	21
6.4.3 Data Flow Diagram	22
<b>6.5 Student_Pass_Percentage</b>	<b>23</b>
6.5.1 Source	23
6.5.2 Target	23
6.5.3 Data Flow Diagram	23
<b>CHAPTER 7</b>	<b>25</b>
<b>CONCLUSION</b>	<b>25</b>
<b>REFERENCES</b>	<b>26</b>
<b>LINKS</b>	<b>27</b>

## LIST OF ACRONYMS AND ABBREVIATION

Acronym/Technical Term	Description
ETCL	Extract Transform Cleansing and Loading
SRC	Source
JNR	Joiner
TGT	Target
LKP	Lookup
WF	Workflow
EXP	Expression
ETL	Extract, Transform, Load
SRT	Sorter
AGG	Aggregate

## LIST OF FIGURES

<b>S. No.</b>	<b>Figures</b>	<b>Page Number</b>
1	Figure 1.1 Schematic Representation of ETL	1
2	Figure 6.1 Flow diagram for Customer Table	17
3	Figure 6.2 Flow diagram for Student Table	18
4	Figure 6.3 Flow diagram for Marks Fact Table	20
5	Figure 6.4 Flow diagram for Subject Average Aggregate Table	22
6	Figure 6.5 Flow diagram for Student Pass Percentage Table	23



## LIST OF TABLES

S. No	Table	Page Number
1	Table 1.1: Some of the commonly used transformations in Informatica	6
2	Table 2.1: Data Source and access mechanisms	8
3	Table 4.1: Department Dimension	11
4	Table 4.2: Student Dimension	11
5	Table 4.3: Marks Fact Dimension	12
6	Table 4.4: Subject Aggregate Dimension	13
7	Table 4.5: Subject Pass Percentage Aggregate Dimension	14
8	Table 5.1: ETL Design Overview	15
9	Table 5.2: Informatica Components	16
10	Table 6.1: Transformation and Justification for Customer Table	17
11	Table 6.2: Transformation and Justification for Student Table	18
12	Table 6.3: Transformation and Justification for Student Table	20
13	Table 6.4: Transformation and Justification for Subject Average Aggregate Table	22
14	Table 6.5: Transformation and Justification for Student Pass Percentage Table	24

## **ABSTRACT**

The purpose of the software requirements document is to systematically capture requirements for the project and the system “Student report” to be developed. Functional requirements of this system are captured in this document. It also serves as the input for the project scoping.

The motivation behind the undertaking is to assemble an information stockroom explicit to business necessity utilizing Informatica Power Center ETL instrument. Understudy data framework is battling the day-by-day record so they chose to manufacture a framework for following the understudy information.

In the Student Report System, we are intending to an assemble a Dataware House for the organization to improve their choice help. The College organization is examining dependent on focal archive of information gathered and is report dependent on the equivalent.

# CHAPTER 1

## INTRODUCTION

The purpose of the project is to build a data warehouse specific to business requirement using Informatica Power Center ETL tool. Student information system is struggling the daily record so they decided to build a system for tracking the student data.

### 1.1 ETL

In this project, the process we are using is ETL. ETL stands for extract, transform, load. Three database function are used in combination to make one tool to take out data from one database and place it into another database.

**Extract:** It is the process of reading the data from database. This is the first stage; the data is taken out from multiple and different type of databases.

**Transform:** It is the process of converting the extracted data from previous state into the state it needs to be so that it can be placed in another database. Transformation occurs by combining data with other data, lookup table, expression etc.

**Load:** It is the process of writing the data into target database.

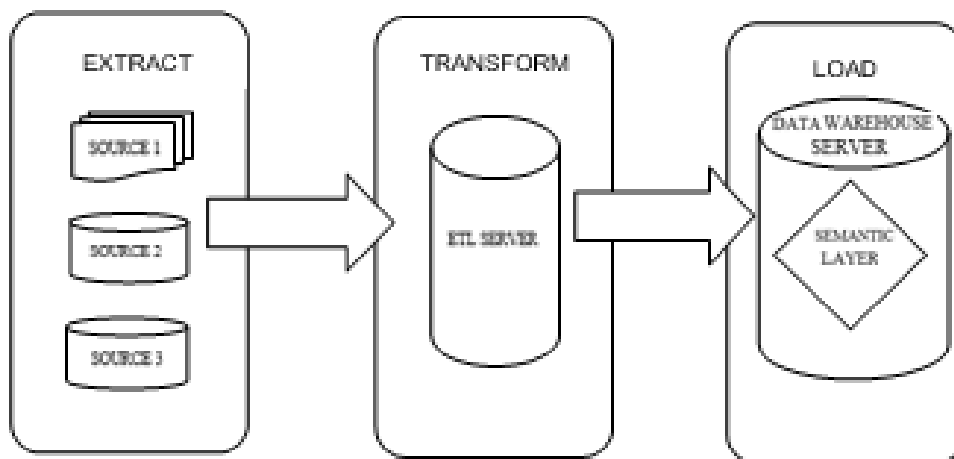


Figure 1.1 Schematic Representation of ETL

### **1.1.1 How ETL Works?**

Data from one or more sources is extracted and then copied to the data warehouse. When dealing with massive data and multiple source systems, the data is consolidated. ETL is used to migrate data from one database to another, and is commonly the precise method needed to load information to and from DataMart and Data Warehouses, but is a process that is also used to transform databases from one format or type to another.

### **1.1.2 Informatica Tool**

Informatica Power center ETL/Data Integration tool is a most widely used tool offering the potential to connect & fetch data from different heterogeneous source and processing of data. Informatica comes to the picture when data system available and at the backend, need to perform certain operations on the data. It can be like cleaning up of data, modifying the data, etc. based on certain set of rules or simply loading of bulk data from one system to another. Informatica offers a rich set of features like operations at row level on data, integration of data from multiple structured, semi-structured or unstructured systems, scheduling of data operation. It also has the feature of metadata, so the information about the process and data operations are also preserved.

### **1.1.3 Architecture**

The architecture of Informatica Power Center is based on the Service Oriented Architecture (SOA) concept. A service-oriented architecture (SOA) can be defined as a group of services, which communicate with each other. The process of communication involves either simple data transfer or it could involve two or more services coordinating same activity.

Therefore, the Informatica Power Center tool consists of two components. They are:

Client component

Server component

### **1.1.4 Client Components in Informatica**

The Power Center Client consists of applications that developers use to design mappings, mapplets, create sessions and workflows to load the data, and monitor workflow progress. In this article, let us discuss about the following applications and tools, which is used by developers during the development process.

Power Center Designer

Workflow Manager

Workflow Monitor

Repository Manager

#### **Power Center Designer**

The Designer has tools to assist you build mappings and mapplets therefore you'll specify the way to move and rework information between sources and targets. The Designer helps you produce supply definitions, target definitions, and transformations to create the mappings. The Designer helps you to work with multiple tools at just one occasion and to work in multiple folders and repositories at an equivalent time. It additionally includes windows therefore you'll read folders, repository objects, and tasks. you'll put together general Designer settings, like background color and font. you'll additionally put together specific tool settings for every Designer tool.

#### **Workflow Manager**

In progress Manager, you outline a group of instruction referred to as a workflow to execute mapping you build in the Designer. Generally, a workflow contains a session and other task you'll wish to perform after you run a session. Task will embrace a session, email notification, or scheduling info. You connect every task with links within the progress. You'll conjointly produce a Worklet within the Workflow Manager. A worklet is an object groups a set of tasks. A worklet is comparable to a Workflow, but without scheduling information. You can run a batch of Worklets inside a Workflow. After you a create a workflow, you run the Workflow in the

Workflow Manager and monitor on the Workflow Monitor.

### **Workflow Monitor**

It is possible to monitor workflows and tasks in the Workflow Monitor. With the workflow Monitor, you'll be able to read details a couple of workflow or task in Gantt chart read or Task read. The workflow Monitor displays workflows that have run a minimum of once. you'll be able to run, stop, abort, and resume workflows from the workflow Monitor. The Workflow Monitor continuously receives information from the Integration Service and Repository Service. It conjointly fetches data from the repository to show historic data.

### **Repository Manager**

This is more of an object management or administrative tool, where the developers spent the least amount of time. You can navigate through multiple folders and repositories and perform basic repository tasks like moving mappings, sessions, workflows or the other objects between folders or repositories with the Repository Manager. Menu items in the Repository Manager are enabled or disabled according to the privileges and permissions you have.

#### **1.1.5 Sources & Targets**

Informatica being an ETL and Data integration tool, you would be always handling and transforming some form of data. The input to our mappings in Informatica is called source system. We import source definitions from the source and then connect to it to fetch the source data in our mappings. There can be different types of sources and can be located at multiple locations. Based upon your requirement the target system can be a relational or flat file system. Flat file targets are generated on the Informatica server machine, which can be transferred later on using ftp.

### **Relational**

These types of sources are database system tables. These database systems are generally owned by other applications which create and maintain this data. It can be a Customer Relationship Management Database, Human Resource Database, etc. for

using such sources in Informatica we tend to either get a duplicate of those datasets, or we tend to get choose privileges on these systems.

## **Flat Files**

Flat files are most common data sources after relational databases in Informatica. A flat file can be a comma separated file, a tab delimited file or fixed width file. Informatica supports any of the code pages like ASCII or Unicode. To use the flat file in Informatica, its definitions must be imported similar to as we do for relational tables.

### **1.1.6 Informatica Transformation Overview**

A transformation is an object that reads, modifies and passes the data on. It represents the set operations performed on the data. It can be categorized in two classes- Active/Passive or Connected/Unconnected.

#### **Active Transformation**

With the help of Active Transformation, we can alter the no. of rows which is passes through the transformation and can alter the row type or transaction boundary. For example, Filter, Transaction Control and Update Strategy are active transformations.

The following are the list of active transformations used for processing the data –

- Source quilter transformation
- Filter transformation
- Ruler transformation
- Rank Transformation
- Sorter transformation
- Joiner transformation
- Union Transformation
- Aggregate Transformation
- Transaction control transformation
- Normalize transformation
- Update strategy transformation
- SQL Transformation

#### **Passive Transformation**

With the help of passive transformation, we cannot alter the no. of rows which goes through it and maintains the row type and transaction boundary.

The following are the list of passive transformations used for processing data.

- Expression transformation
- Sequence generated transformation
  - Stored procedure transformation
  - Look up transformation
  - XML source qualifier transformation
  - SQL Transformation

Transformations can be Connected or Unconnected to the data flow.

### **Connected transformation**

Connected transformation is linked to other transformations or directly to destination table in the mapping.

### **Unconnected transformation**

Unconnected transformation is not linked to other transformations in the mapping. It is invoked within a different transformation and gives a value to that transformation.

**Table 1.1:** Some of the commonly used transformations in Informatica are

<b>Transformation</b>	<b>Description</b>
Aggregator	This transformation is used to perform aggregate calculations such as averages and sums. It performs calculation on row-by-row basis.
Expression	This transformation calculates the value in a single row and tests the conditional statements
Filter	It filters the rows in a mapping. All the ports are input/output in nature and rows that meet the filter conditions pass through it.
Joiner	This transformation joins two heterogeneous sources.
Lookup	It searches the data in the relational table and returns it. A user can use multiple lookups at a time.

## **1.2 Scope**

In the Student Report System, we are planning to a build a Data ware



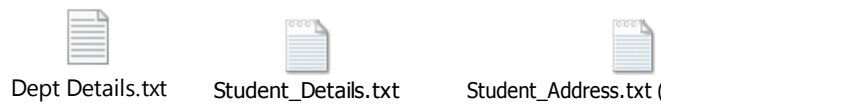
House for the administration to enhance their decision support. The College administration is analyzing based on central repository of data collected and is report based on the same. Department, Student dimension and Marks Fact are included. Subject Average Aggregate and Subject Pass Percentage Aggregate to produce ETL process.

### 1.3 Assumptions, Dependencies & Constraints

#### General constraints and Assumptions:

- Repository is connected and running. Oracle SQL Developer is connected to database.

- Source files are available in the particular path which is



## CHAPTER 2

### DATA SOURCES AND ACCESS MECHANISMS

**Table 2.1:** Data Source and access mechanisms

<b>Data Source Name</b>	<b>Data Source Type</b>	<b>Access Method</b>
Dept Details	Flat file	File reader
Student marks	Flat file	File reader
Student_Details	Flat file	File reader
Student_Address	Flat file	File reader
Dept_Dimension	Database	Relational reader
Student_Dimension	Database	Relational reader
Marks_Fact	Database	Relational reader
Subject_Average_Aggregate	Database	Relational reader
Subject_Pass_Percent	Database	Relational reader

There are four flat files(.txt) as the input to the ETL software which are Student Marks.txt, Dept Details.txt, Student Details.txt, Student Address.txt. These files act as source. All the raw data is stored in these files. Data is extracted from these files using file reader method to get a fruitful output.

There are five database files as the output to the ETL software which are Dept\_Dimension, Student\_Dimension, Marks\_Fact, Subject\_Average\_Aggregate and Subject\_Pass\_Percent. In these files the transformed data will be stored using relational reader method.

## CHAPTER 3

### SOURCE SYSTEM CLEANING RULES

#### 3.1 Department Dimension

- Dept\_code should be a valid varchar (3) datatype.
- Dept\_Name should be a valid datatype.

#### 3.2 Student Dimension

- Student\_Id should be a valid datatype.
- Student\_Name should be a valid datatype varchar (20).
- Postal code column should be cleaned to remove any extra space and special characters.
- Phone Number column should be a valid 10-digit phone number, if there are any invalid number found then phone number has to be loaded as NULL.
- All characters in address column has to be convert to upper case.

#### 3.3 Marks Fact Dimension

- Student\_ID should be of the valid number (3).
- Student\_Name should be of the valid varchar.
- Dept\_ID should be of the valid number (3)
- Dept\_Name should be of the valid varchar
- Total should be sum of the marks of the subject
- Average should be average of the marks of each subject

#### 3.4 Subject Aggregate Dimension

- Dept\_ID should be of the valid number (3).
- Dept\_Name should be of the valid varchar
- Subject1\_Average is the average of the marks of subject1 scored by students
- Subject2\_Average is the average of the marks of subject2 scored by students

- Subject3\_Average is the average of the marks of subject3 scored by students
- Subject4\_Average is the average of the marks of subject4 scored by students
- Subject5\_Average is the average of the marks of subject5 scored by students

### **3.5 Subject\_Pass\_Percentage Dimension**

- Dept\_ID should be of the valid number (3)
- Dept\_Name should be of the valid varchar
- We should check the Pass Percentage for each student with respect to each subject.

## CHAPTER 4

### TARGET SOURCE MAPPING AND TRANSFORMATION RULES

#### 4.1 Department Dimensions

Table 4.1: Department Dimension

Source column names	Mapping logic	Target column names
DEPT_CODE	DEPT_ID Should be of the valid varchar (3) or else abort the session	DEPT_ID
DEPT_NAME	We need to convert the DEPT_NAME into upper case.	DEPT_NAME

#### 4.2 Student Dimension

Table 4.2: Student Dimension

Source Column Name	Mapping Logic	Target Column Name
STUDENT_ID	STUDENT_ID should be of the valid varchar (3) or else abort the session	STUDENT_ID
STUDENT_NAME	STUDENT_NAME should be of varchar (20)	STUDENT_NAME
DEPT_ID	DEPT_ID Should be of the valid varchar (3) or else abort	DEPT_ID
ADDRESS	Directly moved to the target	ADDRESS
POSTAL_CODE	Directly moved to the target	POSTAL_CODE

PHONE_NUMBER	Directly moved to the target	PHONE_NUMBER
--------------	------------------------------	--------------

### 4.3 Marks Fact Dimension

**Table 4.3:** Marks Fact Dimension

Source column name	Mapping logic	Target column name
STUDENT_ID	Data directly moved to target table	STUDENT_ID
STUDENT_NAME	Data directly moved to target table	STUDENT_NAME
DEPT_ID	Data directly moved to target table	DEPT_ID
DEPT_NAME	Data directly moved to target table	DEPT_NAME
MARK1	Data directly moved to target table	MARK1
MARK2	Data directly moved to target table	MARK2
MARK3	Data directly moved to target table	MARK3
MARK4	Data directly moved to target table	MARK4
MARK5	Data directly moved to target table	MARK5
TOTAL MARKS	Sum of the marks attained by Student in each subject	O_TOTAL
AVERAGE MARKS	Average of the marks is attained using dividing Total marks by Number of subjects	O_AVERAGE

RESULT	The result is evaluated checking if the resulted average is greater than 50	O_RESULT
--------	---	----------

## 4.4 Subject Aggregate Dimension

**Table 4.4:** Subject Aggregate Dimension

Source column name	Mapping logic	Target column name
DEPT_ID	DEPT_ID should be of the valid varchar (3) or else abort	DEPT_ID
DEPT_NAME	Directly moved to the target	DEPT_NAME
SUBJECT1_PASS_PERCENTAGE	Check whether the SUBJECT1_PASS_PERCENTAGE is greater than 50% or less	SUBJECT1_PASS_PERCENTAGE
SUBJECT2_PASS_PERCENTAGE	Check whether the SUBJECT2_PASS_PERCENTAGE is greater than 50% or less	SUBJECT2_PASS_PERCENTAGE
SUBJECT3_PASS_PERCENTAGE	Check whether the SUBJECT3_PASS_PERCENTAGE is greater than 50% or less	SUBJECT3_PASS_PERCENTAGE
SUBJECT4_PASS_PERCENTAGE	Check whether the SUBJECT4_PASS_PERCENTAGE is greater than 50% or less	SUBJECT4_PASS_PERCENTAGE
SUBJECT5_PASS_PERCENTAGE	Check whether the SUBJECT5_PASS_PERCENTAGE is greater than 50% or less	SUBJECT5_PASS_PERCENTAGE

## 4.5 Subject\_Pass\_Percentage\_Aggregate Dimension

**Table 4.5:** Subject Pass Percentage Aggregate Dimension

Source column name	Mapping logic	Target column name
DEPT_ID	DEPT_ID should be of the valid varchar (3) or else abort	DEPT_ID
DEPT_NAME	Directly moved to the target	DEPT_NAME
SUBJECT1_AVERAGE	We need to average the total marks of the students scored in the subject1 using respective function	SUBJECT1_AVERAGE
SUBJECT2_AVERAGE	We need to average the total marks of the students scored in the subject2 using respective function	SUBJECT2_AVERAGE
SUBJECT3_AVERAGE	We need to average the total marks of the students scored in the subject3 using respective function	SUBJECT3_AVERAGE
SUBJECT4_AVERAGE	We need to average the total marks of the students scored in the subject4 using respective function	SUBJECT4_AVERAGE
SUBJECT5_AVERAGE	We need to average the total marks of the students scored in the subject5 using respective function	SUBJECT5_AVERAGE



## CHAPTER 5

### ETL DESIGN OVERVIEW AND COMPONENTS

**Table 5.1:** ETL Design Overview

Module ID	Description	Technology	Document Reference	Module
1	To load department data this module is used	INFORMATICA	DW – Building ETL Processes  Software Requirements Document	Department
2	To load student data this module is used			Student
3	To load marks data this module is used			Marks
4	To load subject average data this module is used			Subject Average
5	To load Subject Pass Percentage data this module is used			Subject Pass Percentage

**Table 5.2:** Informatica Components

Workflow Sequence		Mapping name	Session name	Work flow name
1	Department	m_Department	s_m_Department	wf_Student_Report
2	Student	m_Student	s_m_Student	
3	Marks Fact	m_marks_fact	s_m_marks_fact	
4	Subject_Average_Aggregate	m_subject_average	s_m_subject_average	
5	Subject_Pass_Percentage	m_subject_pass_percentage	s_m_subject_pass_percentage	

Informatica is used as a technology to complete are all the modules that are Department, Student, Marks, Subject Average, Subject Pass Percentage. All the Document Reference were provided in the Software Requirement Document and DW – Building ETL Process

There are total five mapping in the project which are m\_Department, m\_Student, m\_marks\_fact, m\_subject\_average, m\_subject\_pass\_percentage where all the transformations are made are to the requirement given by the business unit

Each mapping has its own session layer which are s\_m\_Department, s\_m\_Student, s\_m\_marks\_fact, s\_m\_subject\_average, s\_m\_subject\_pass\_percentage These session layers are sequenced together according to data flow in a single workflow which is wf\_Student\_Report. So that all the files are loaded in a specific order and dependent once are loaded first.

# CHAPTER 6

## ETL PROGRAM SPECIFICATIONS

### 6.1 Customer Dimension

#### 6.1.1 Source

Source file is flat file, tab delimited (<tab>).

**Customer File Name:** Dept Details.txt

#### 6.1.2 Target

**Customer Table:** Department\_Dimension

#### 6.1.3 Data Flow Diagram

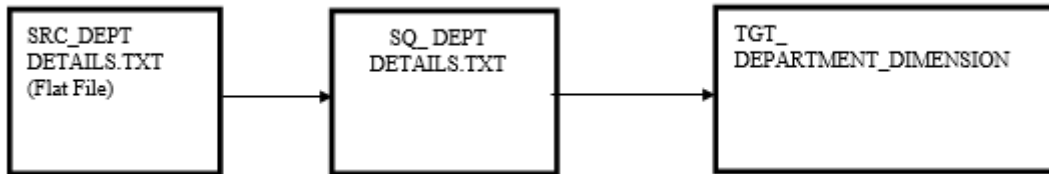


Figure 6.1 Flow diagram for Customer Table

Table 6.1: Transformation and Justification for Customer Table

Transformations	Justifications
SRC_DEPT DETAILS.TXT	This is a source definition for department source file.
SQ_DEPT DETAILS.TXT	This is a source qualifier for department source file. No overwrite.  The data is pulled directly.
TGT_ DEPARTMENT_DIMENSION	It is a target definition for department dimension.

## 6.2 Student Dimension

### 6.2.1 Source

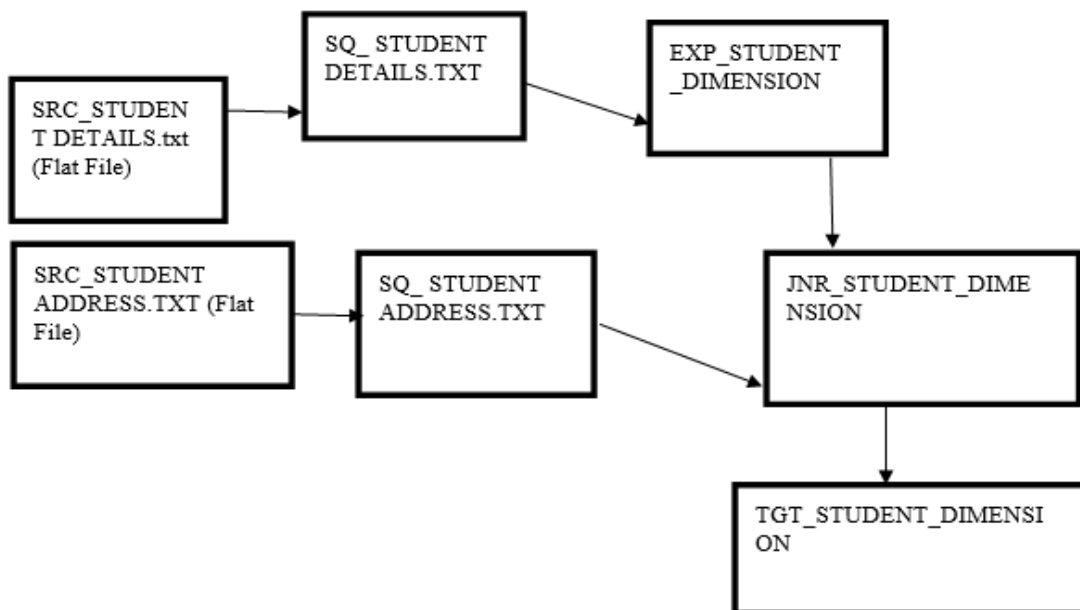
Source file is flat file, tab delimited (<tab>).

**Supplier File Name:** STUDENT\_DETAILS.TXT and STUDENT\_ADDRESS.TXT.

### 6.2.2 Target

**Supplier Table:** STUDENT DIMENSION

### 6.2.3 Data Flow Diagram



**Figure 6.2** Flow diagram for Student Table

**Table 6.2:** Transformation and Justification for Student Table

Transformations	Justifications
SRC_STUDENT_DETAILS.TXT	This is the source definition for student source file.
SQ_STUDENT_DETAILS.TXT	This is the source qualifier for student source. No override. Data is pulled directly.

JNR_STUDENT_DIMENSION	This is joiner transformation is used to connect Student Details and Student Address..
EXP_STUDENT_DIMENSION	This expression transformation is used to remove spaces and special character in Postal code, to check Phone Number is valid and is 10 digits. if there are any invalid number found then phone number has to be loaded as NULL
TGT_STUDENT_DIMENSION	This is the target definition for student source file used to insert.

### **6.3 Marks\_Fact**

#### **6.3.1 Source**

Source files are flat file and dimension table, tab delimited (<tab>).

**Product File Name:** STUDENT\_MARKS.TXT and TGT\_STUDENT DIMENSION

#### **6.3.2 Target**

**Product Target Table:** TGT\_MARKS\_FACT

### 6.3.3 Data Flow Diagram

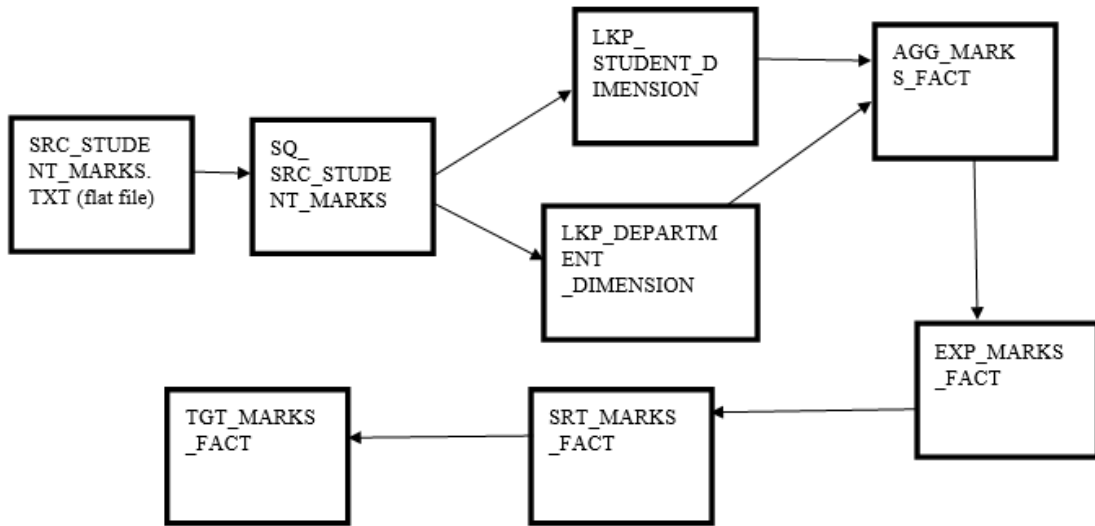


Figure 6.3 Flow diagram for Marks Fact Table

Table 6.3: Transformation and Justification for Student Table

Transformations	Justifications
SRC_STUDENT_MARKS.TXT	This is a source definition for marks fact table.
SQ_SRC_STUDENT_MARKS	This is a source qualifier for student marks source file. No overwrite. The data is pulled directly.
LKP_STUDENT_DIMENSION	It is used to lookup the student dimension based on Dept_id of the department and loads the corresponding student dimension.
LKP_DEPARTMENT_DIMENSION	It is used to lookup the student dimension based on Dept_id of the department and loads the corresponding department dimension.

AGG_MARKS_FACT	This aggregator transformation is used to find the sum of marks of subjects and the average of the subjects and highest score of the subjects.
EXP_MARKS_FACT	It is an expression transformation used to tell whether the student is pass or fail based on his marks  Student is pass if he scores above or equal to 50% and fail if he scores below 50% in all subjects
SRT_MARKS_FACT	This sorter transformation is used to sort the result based on Student_ID.
TGT_MARKS_FACT	It is a target definition for marks fact table.

## 6.4 Subject\_Average\_Aggregate

### 6.4.1 Source

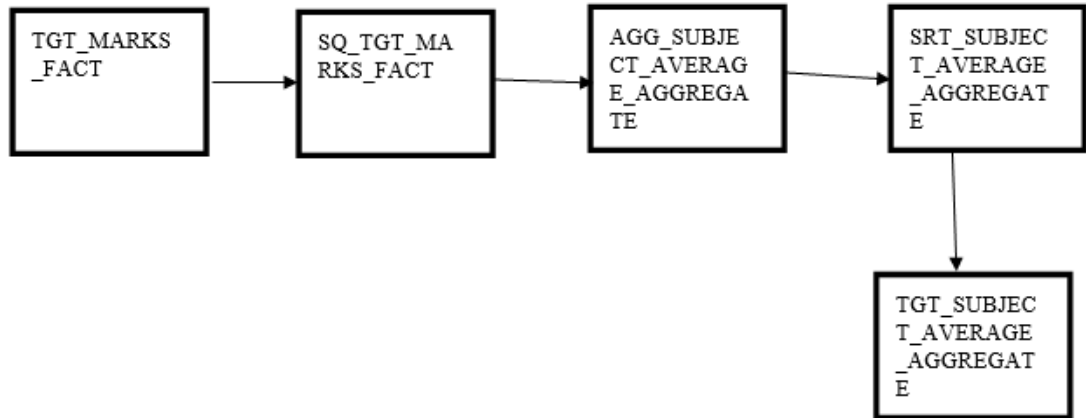
Source dimension is MARKS\_FACT, tab delimited (<tab>).

**Aggregate Source Name:** TGT\_MARKS\_FACT

### 6.4.2 Target

**Aggregate Target Table:** TGT\_STUDENT\_AVERAGE\_AGGREGATE

### 6.4.3 Data Flow Diagram



**Figure 6.4** Flow diagram for Subject Average Aggregate Table

**Table 6.4:** Transformation and Justification for Subject Average Aggregate Table

Transformations	Justifications
TGT_MARKS_FACT	This is the source definition for subject average aggregate table
SQ_TGT_MARKS_FACT	This is the source qualifier for subject average aggregate table. No override. Data is pulled directly.
AGG_SUBJECT_AVERAGE_AGGREGATE	This aggregator transformation is use to find the average of each subject.
SRT_SUBJECT_AVERAGE_AGGREGATE	This sorter transformation is use to sort result based on Dept_ID.
TGT_SUBJECT_AVERAGE_AGGREGATE	This is the target definition for subject average aggregate table



## 6.5 Student\_Pass\_Percentage

### 6.5.1 Source

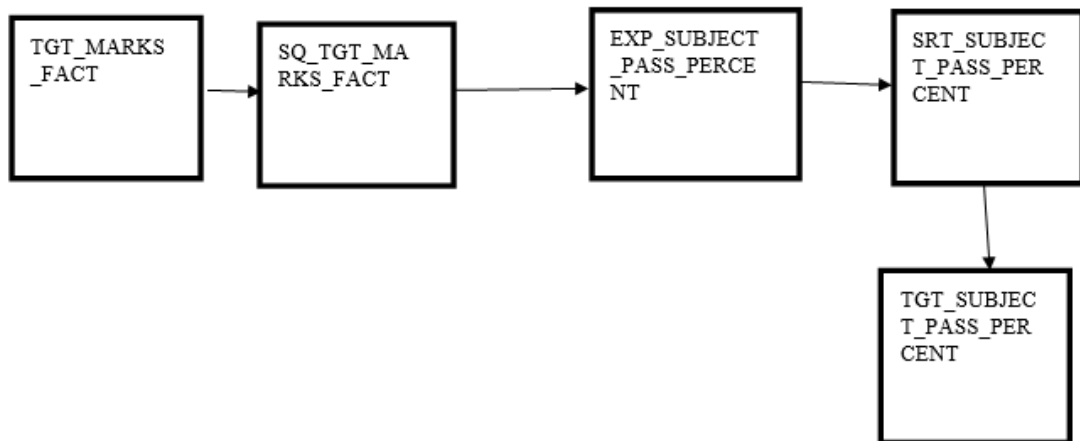
Source file is a flat file.

**Time File Name:** TGT\_MARKS\_FACT

### 6.5.2 Target

**Time Target Table/File Name:**  
TGT\_SUBJECT\_PASS\_PERCENTAGE\_AGGREGATE

### 6.5.3 Data Flow Diagram



**Figure 6.5** Flow diagram for Student Pass Percentage Table

**Table 6.5:** Transformation and Justification for Student Pass Percentage Table

<b>Transformations</b>	<b>Justifications</b>
TGT_MARKS_FACT	This is a source definition for subject pass percentage aggregate table.
SQ_TGT_MARKS_FACT	This is a source qualifier for subject pass percentage table. No overwrite. The data is pulled directly.
EXP_SUBJECT_PASS_PERCENTAGE_AGGREGATE	This expression transformation is used to calculate pass percentage for each subject.
SRT_SUBJECT_PASS_PERCENT	This sorter transformation is use to sort result based on Dept_ID.
TGT_SUBJECT_PASS_PERCENTAGE_AGGREGATE	It is a target definition for pass percentage aggregate table dimension.

## CHAPTER 7

### CONCLUSION

Informatica Powercenter is one among the leading ETL tool presently accessible in market. Not just for its simplicity to grasp and manage the information flow, however additionally for the options it offers to the user. the globe is driven by Data solely currently. ranging from a pencil to a social unit selling in BigBazaar, the vendor needs to keep track of its item data in and out of its garage (here in ETL perspective, u might say Data Warehouse). When there was not that huge consumer-centric market, the buying and selling equation could be easily possible to track in excel sheets or small transaction databases.

But with the growing capability of selling/buying/consuming/importing/exporting, the user extremely must handle a large quantity of information that can't be handled by a mere document or excel sheet or small transaction database like oracle or mysql. Then came the thought of Data warehousing, data-mining, data-modelling and massive stuffs. however so as to try and do of these things, we want to initial store these Big Data somewhere in no time. With the growing speed generating Data, the necessity of storing them during a real time basis, hit the necessity of the market. that is wherever ETL tools came into play, and Informatica Powercenter is one among the leading product among them. Since, no business can sustain without these data in this highly competitive market, the need for Informatica will always be a hotcake.

The scope of Informatica is already immense & it includes its capability to extract, process & load Data to/from any accessible system, application, website, database, API, IOT, etc.

Tools like Informatica will really perform things that a lot of big data Developers would find complex. Tools like Informatica are well-equipped & improving each single day to accelerate processing for Big Data

## REFERENCES

- [1] "Cognizant to Acquire Mustache, a New York-Based Creative Content Agency". *News | Cognizant Technology Solutions*. Retrieved December 11, 2018.
- [2] Ralph, Kimball, (2004). *The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data*. Caserta, Joe, 1965-. Indianapolis, IN: Wiley.
- [3] Informatica, *Informatica website*, retrieved 15 May 2010
- [4] "What is Agile Software Development?" *Agile Alliance*. 8 June 2013. Retrieved 4 April 2015.

## **LINKS**

Informatica ETL Basics <https://www.guru99.com/introduction-informatica.html>

Informatica ETL Working <https://www.webopedia.com/TERM/E/ETL.html>

Informatica Power Center <https://www.edureka.co/blog/what-is-informatica/>

PLAGIARISM VERIFICATION REPORT

Date: 21/05/2019

Type of Document (Tick):  PhD Thesis  M.Tech Dissertation/ Report  B.Tech Project Report  Paper

Name: Deepankar Mathur Department: ECE Enrolment No 151064

Contact No. 7597960002 E-mail: dmathur09@gmail.com

Name of the Supervisor: Dr Meenakshi Sood

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): STUDENT REPORT

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/ revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

Complete Thesis/Report Pages Detail:

- Total No. of Pages = 37
- Total No. of Preliminary pages = 9
- Total No. of pages accommodate bibliography/references = 2

Deepankar Mathur  
(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found Similarity Index at 28 (%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

Hellish  
21/5/19  
(Signature of Guide/Supervisor)

[Signature]  
Signature of HOD

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Generated Plagiarism Report Details (Title, Abstract & Chapters)	
			Word Counts	
<u>22.05.2019</u>	<ul style="list-style-type: none"> <li>• All Preliminary Pages</li> <li>• Bibliography/Images/Quotes</li> <li>• 14 Words String</li> </ul>	<u>26 %</u>	Word Counts	<u>3,804</u>
Report Generated on			Character Counts	<u>21,840</u>
<u>20.05.2019</u>		Submission ID	Total Pages Scanned	<u>27</u>
		<u>1134307466</u>	File Size	<u>852 K</u>

[Signature]  
Checked by  
Name & Signature

Ashok

[Signature]  
Librarian 22/05/2019

Please send your complete thesis/report in (PDF) with Title Page, Abstract and Chapters in (Word File) through the supervisor at [plagcheck.juit@gmail.com](mailto:plagcheck.juit@gmail.com)