7.1 Software Development Methodology

The Spiral Modeling approach is used in this study. Steps of the development methodology are summarized in Figure 8.1. The focus is primarily on data modeling requirements. Spiral development method has been evolving based on experience with various refinements of the waterfall model as applied to software projects. The Spiral model reduces risk early by breaking down the project into smaller iterations. This causes part of the testing and design to occur earlier in the project, which in turn allows us more time to fix problems as they are discovered. This increases the probability that either the project will succeed or we will realize early that we cannot complete the project successfully within the allocated time and budget. The starting point is to perform specific business data analysis process to identifying the DW project by a group of end user through solution integration all steps will be performed. Figure 7.1 shows Spiral Development Method cycle.

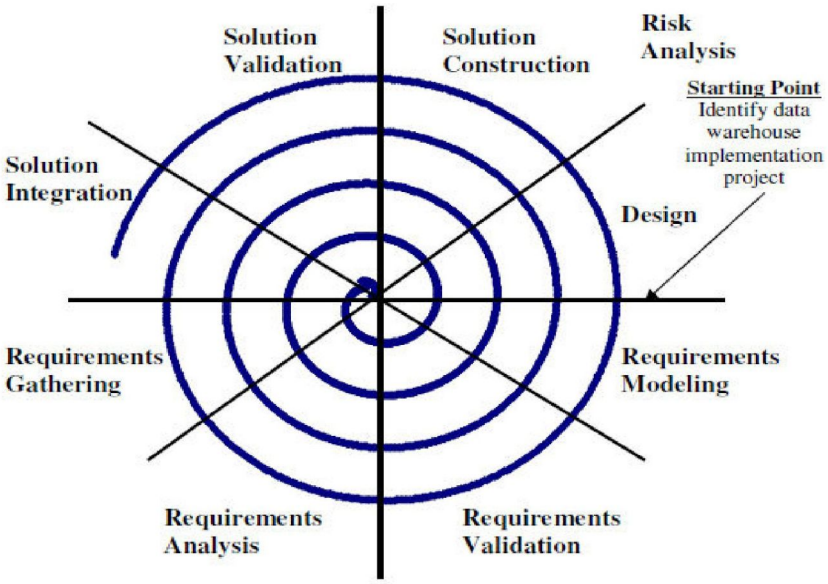


Figure 7.1: Spiral Development Method Cycle

7.1.1 Requirement Gathering

This phase of the lifecycle is concerned with the understanding of the business needs and data requirements of the users of the system. During requirements gathering, end-user requirements are collected and documented. Requirements gathering is often incorporated in some way into

studies of the business process and information analysis activities in which end users are involved. Requirements gathering therefore are very much oriented toward understanding the problem domain for which the modeling will be done. According to Kimball [6], business users and their requirements impact almost every decision made throughout the implementation of a DW. Gathering data requirements phase includes:

1) Conducting user interviews
2) Documenting results

7.1.2 Requirement Analysis

During requirement analysis, informal end-user requirements are further investigated, and initial dimensional models are produced showing facts, measures, dimension keys, and dimension hierarchies. Dimension hierarchies can include parallel hierarchical paths. Models produced during requirements analysis must be kept simple, because these initial dimensional models must be discussed with end users.

7.1.3 Requirements Validation

Initial dimensional models are used in this process of validating the end-user requirements and for assessing the scope and impact of the development project.

7.1.4 Requirements Modeling

In this phase of the cycle, validated user requirements are modeled. These models are further developed into detailed dimensional models, showing all elements of the model and their properties. Detailed dimensional models can further be extended and optimized.

7.1.5 Risk Analysis

The purpose of the risk analysis phase is to assess both technical and management risks.

7.1.6 Design, Construction, Validation, and Integration

Design and construction activities have to be performed after the requirements are modeled. This step will possibly extend and change the models produced in the previous stages. After that, validation of the proposed system has to be performed with the users of the system. According to user satisfaction the integration part is done. If any changes are needed they are added to the system.

The following table shows some of the fact tables, dimensional tables and primary keys of the fact tables used in SDW.

FACT TABLES	DIMENSION TABLES	FOREIGN KEYS OF THE FACT TABLES
ADMISSION_FACT	admission, department, college, batch	usn, Department Id, College Id, Batch Id
RESULT_FACT	result_cgpa_new, department, college, batch	usn, Department Id, College Id, Batch Id
PROJECT_FACT	projects, subject_info	usn, subject_code

Table 7.1: Tables of the system

Following is the list of the fact tables used in the implementation of SDW:

SUBJECT_FACT
Department Id
Batch Id
Subject Id
Total Credits

ADMISSION_FACT
usn
Department Id
Batch Id
College Id
Total Students

PROJECT_FACT
usn
Department Id
Batch Id
Grade

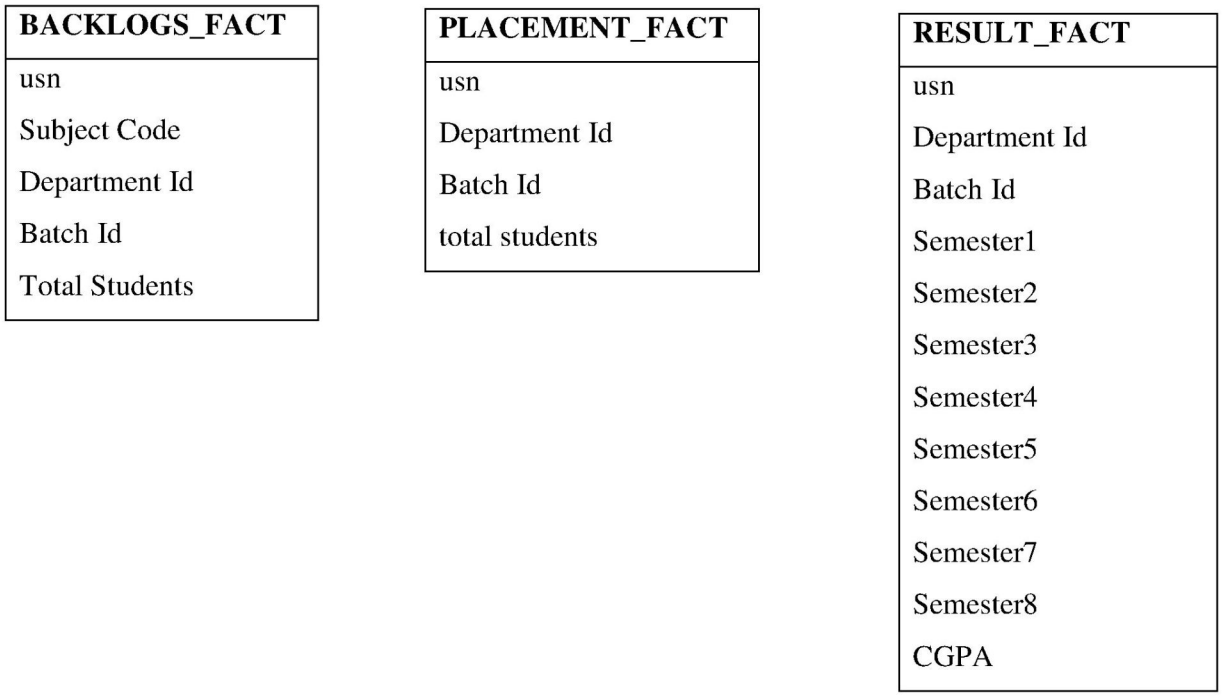
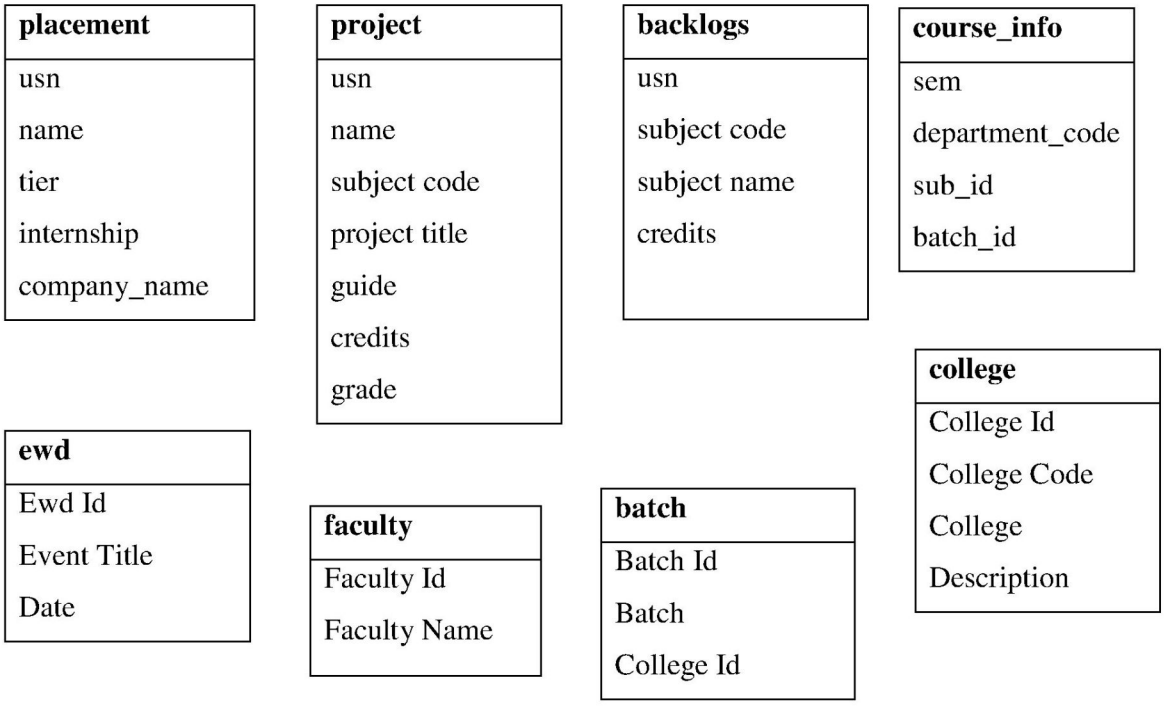


Figure 7.2 Fact tables of SDW

Following is the list of dimension tables used in the implementation of student data warehouse.



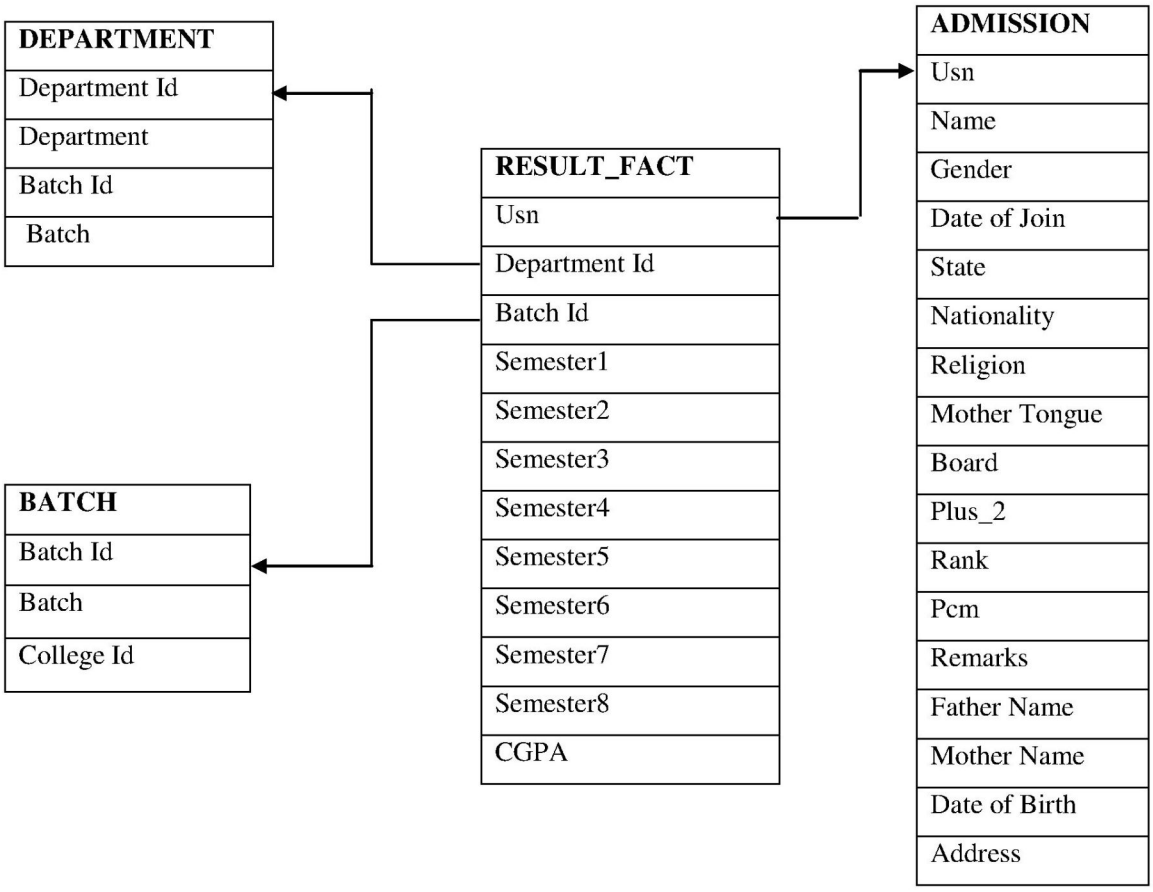


Figure 7.4: Dimensional Model for Admission

The above diagram shows the data model (star schema) for the admission table. In this ADMISSION_FACT is the fact table which contains primary keys of dimension tables. Each fact table is associated with the corresponding cube in the catalog file. BATCH, ADMISSION, DEPARTMENT AND COLLEGE are the dimension tables in the above data model. These dimension tables contains the descripted information about the keys in the fact table. The fact table also contains measures. In the above fact table Total Students is the fact table, which is used to calculate the total number of students admitted to the college/branch in the particular year.

The following is the data model used for the implementation of the Result cube of SDW.

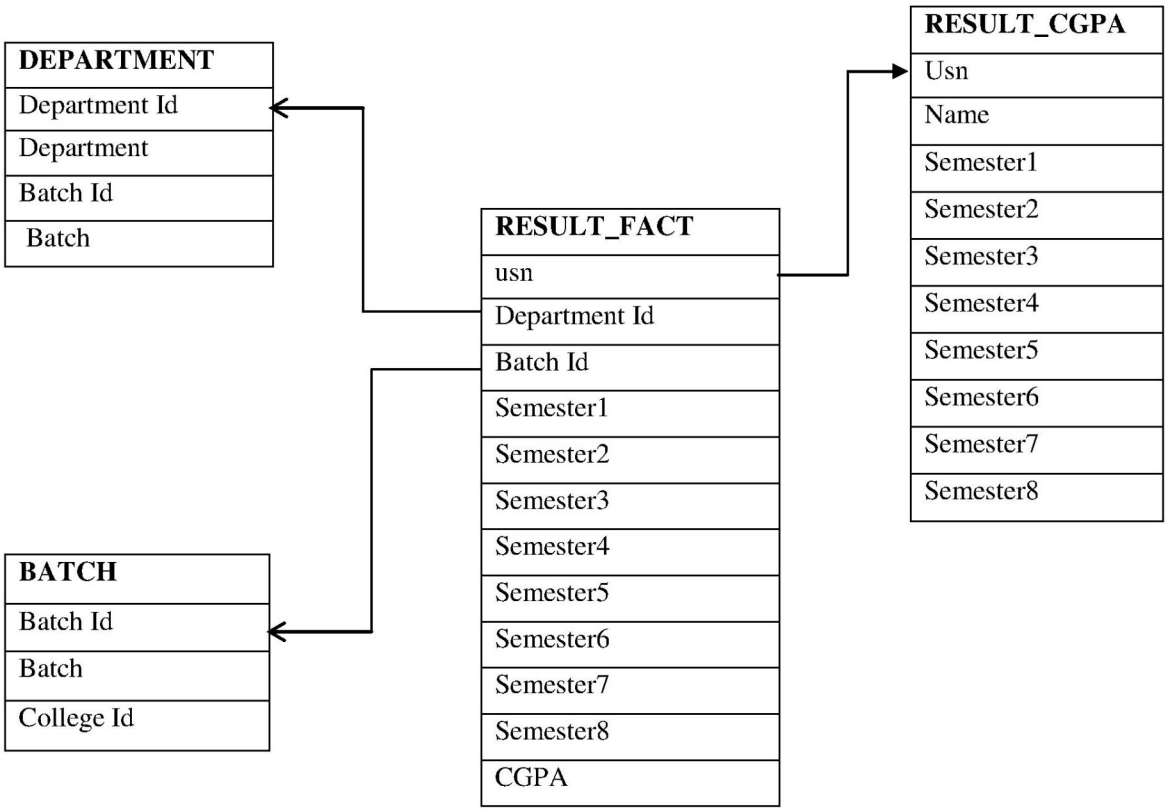


Figure 7.5: Dimensional model for Result

The above figure shows the data model for the Result cube. It has RESULT_FACT as the fact table. It contains usn, Department Id and Batch Id as the foreign keys which refers to the dimension tables RESULT_CGPA, DEPARTMENT and BATCH respectively. The fact table is associated with the cube named "Result" in the catalog file. The fact also contains the Semester1, Semester2, Semester3, Semester4, Semester5, Semester6, Semester7, Semester8 and CGPA as the measures which are used to describe the results of the students.

CHAPTER 8

IMPLEMENTATION

The following is the design of the mondrian schema. Mondrian schema is a catalog file which defines a multidimensional database. It contains a logical model, consisting of cubes, hierarchies, and members, and a mapping of this model onto a physical model. The logical model consists of the constructs used to write queries in MDX language: cubes, dimensions, hierarchies, levels, and members. Mondrian schemas are represented in an XML file. The physical model is the source of the data which is presented through the logical model. It is typically a star schema, which is a set of tables in a relational database. The following section shows the part of the catalog file used for the implementation of the student data warehouse.

sdw.xml.

```
<Schema name="sdw">

<Dimension type="StandardDimension" visible="true" highCardinality="false"
name="Department">
<Hierarchy visible="true" hasAll="true" allMemberName="All Departments"
primaryKey="Department Id">
<Table name="department">
</Table>
<Level name="Department" visible="true" column="Department Code" type="String"
uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>

<Cube name="Admission" visible="true" cache="true" enabled="true">
<Table name="admission_fact">
</Table>
<Dimension type="StandardDimension" visible="true" foreignKey="usn" highCardinality="false"
name="Gender">
<Hierarchy visible="true" hasAll="true" allMemberName="All Gender" primaryKey="usn">
<Table name="admission">
</Table>
```

<Level name="Gender" visible="true" column="Gender" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="usn" highCardinality="false" name="Geography">
<Hierarchy visible="true" hasAll="true" allMemberName="All Region" primaryKey="usn">
<Table name="admission">
</Table>
<Level name="Nationality" visible="true" column="Nationality" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
</Level>
<Level name="State" visible="true" column="State" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="usn" highCardinality="false" name="Board">
<Hierarchy visible="true" hasAll="true" allMemberName="All Board" primaryKey="usn">
<Table name="admission">
</Table>
<Level name="Board" visible="true" column="Board" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="usn" highCardinality="false" name="USN">
<Hierarchy visible="true" hasAll="true" allMemberName="All USN" primaryKey="usn">
<Table name="admission">
</Table>

```
<Level name="USN" visible="true" column="usn" type="String" uniqueMembers="true"
levelType="Regular" hideMemberIf="Never">
<Property name="Name" column="Name" type="String">
</Property>
<Property name="Father Name" column="Father Name" type="String">
</Property>
<Property name="Mother Name" column="Mother Name" type="String">
</Property>
<Property name="Date Of Birth" column="Date of birth" type="Date">
</Property>
<Property name="Date Of Joining" column="Date of join" type="Date">
</Property>
<Property name="PCM Marks" column="pcm" type="Numeric">
</Property>
<Property name="Plus 2" column="Plus_2" type="Numeric">
</Property>
</Level>
</Hierarchy>

<Dimension type="StandardDimension" visible="true" foreignKey="usn" highCardinality="false"
name="Name">
<Hierarchy visible="true" hasAll="true" allMemberName="All Names" primaryKey="usn">
<Table name="admission">
</Table>
<Level name="Name" visible="true" column="Name" type="String" uniqueMembers="true"
levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="usn" highCardinality="false"
name="Father Name">
<Hierarchy visible="true" hasAll="true" primaryKey="usn">
<Table name="admission">
</Table>
```

```
<Level name="Father Name" visible="true" column="Father Name" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="usn" highCardinality="false"
name="Mother Name">
<Hierarchy visible="true" hasAll="true" primaryKey="usn">
<Table name="admission">
</Table>
<Level name="Mother Name" visible="true" column="Mother Name" type="String"
uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="usn" highCardinality="false"
name="Date of Birth">
<Hierarchy visible="true" hasAll="true" primaryKey="usn">
<Table name="admission">
</Table>
<Level name="Date of Birth" visible="true" column="Date of birth" type="Date"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="usn" highCardinality="false"
name="Date of Joining">
<Hierarchy visible="true" hasAll="true" primaryKey="usn">
<Table name="admission">
</Table>
<Level name="Date of Birth" visible="true" column="Date of join" type="Date"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
```

```
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="usn" highCardinality="false"
name="PCM Marks">
<Hierarchy visible="true" hasAll="true" primaryKey="usn">
<Table name="admission">
</Table>
<Level name="PCM Marks" visible="true" column="pcm" type="Numeric"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" foreignKey="usn" highCardinality="false"
name="Plus 2">
<Hierarchy visible="true" hasAll="true" primaryKey="usn">
<Table name="admission">
</Table>
<Level name="Plus 2" visible="true" column="Plus_2" type="Numeric" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>
<DimensionUsage source="Batch" name="Batch" visible="true" foreignKey="Batch Id"
highCardinality="false">
</DimensionUsage>
<DimensionUsage source="Department" name="Department" visible="true"
foreignKey="Department Id" highCardinality="false">
</DimensionUsage>
<Measure name="Total Students" column="usn" datatype="Numeric" aggregator="count"
visible="true">
</Measure>
</Cube>
```

The above catalog file sdw.xml contains declarations for the cubes, dimensions, measures and levels. The cube named “Admission” is used in the catalog file. Dimensions can be either private

or shared. Private dimensions are accessible only inside the cube in which it is declared. In the given catalog file "Name" is the private dimension and it is accessible only inside the Admission cube. The catalog file also contains two shared dimensions: Batch and Department. Shared dimensions are the dimensions which are accessible inside any of the cube. To use the shared dimensions inside a cube we must use <DimensionUsage /> tag. Each cube is associated with the fact table. In the above schema file cube "Admission" is associated with the fact table "admission_fact". The cube contains a measure named "Total Students" which has an aggregator attribute and it is set count. This measure is used to calculate the total number of students admitted to the college in a particular year. The "Department" dimension contains an attribute named "allMemberName" under the <Hierarchy /> element. It is set to the value "All Departments", indicating that admission details of all the departments are included by default.

The following section shows some of the MDX queries which are used to retrieve the values from the database using the catalog file sdw.xml. MDX stands for "multi-dimensional expressions". It is the main query language implemented by Mondrian.

Some of the sample queries used in the implementation of Student Data Warehouse are

1) MDX query for retrieving the admission details based on their department, batch, gender, mother tongue and geographical sections.

```
SELECT {[Measures].[Total Students]} ON COLUMNS,
NON EMPTY Hierarchize ({([Department].[All Departments],[Batch].[All
Batches],[Gender].[All Gender], [Geography].[All Region], [Mother Tongue].[All
Languages], [USN].[All USN])}) ON ROWS
FROM [Admission]
```

2) MDX query for retrieving the result of students based on their departments and batches.

```
SELECT {[Measures].members,[Measures].[CGPA]} ON COLUMNS,
NON EMPTY Hierarchize({([Department].[All Departments],[Batch].[All
Batches],[USN].[All Students])}) ON ROWS
FROM [Result]
```

3) MDX Query for getting the information about the students who got placed in a Tier-1 companies.

```
SELECT {[Measures].[Total Students]} ON COLUMNS,
NON EMPTY Hierarchize ({( [Company Name],[USN].[All USN])}) ON ROWS
FROM [Placement]
WHERE [Tier].[1]
```

8.1 Graphical view of the implementation of SDW:

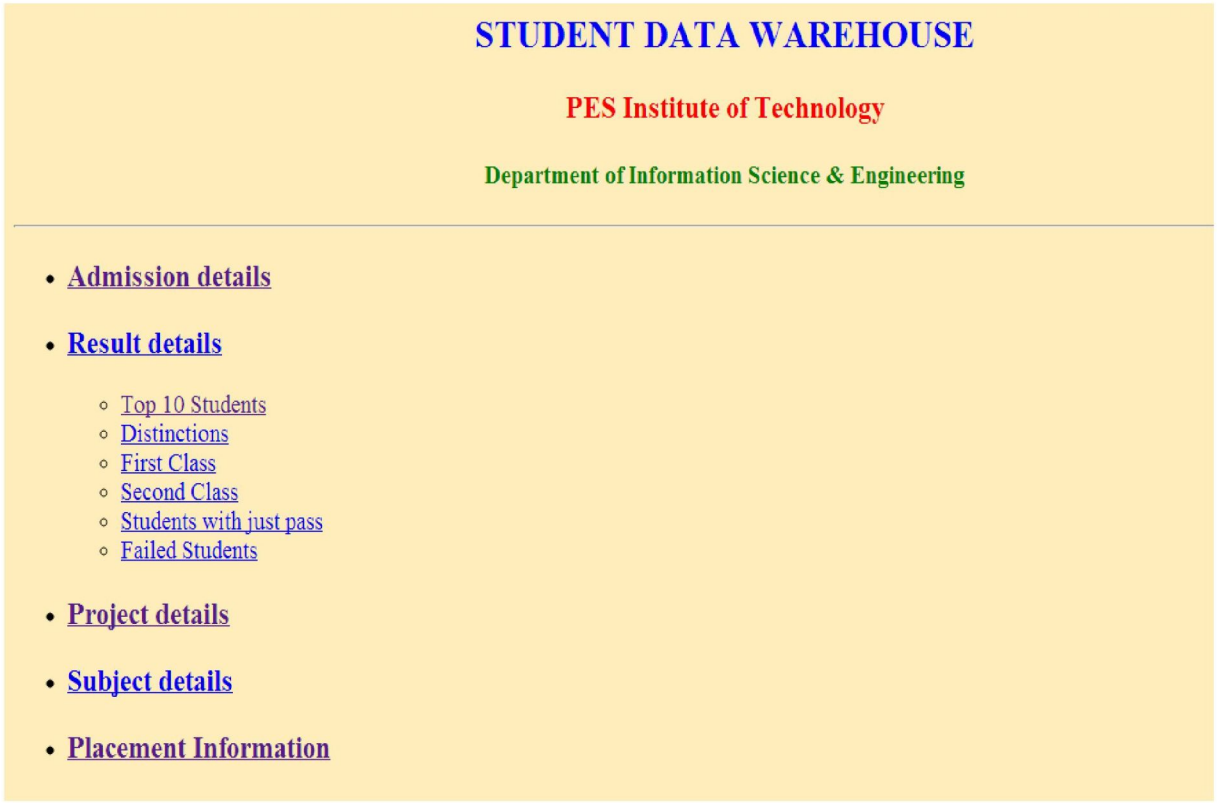
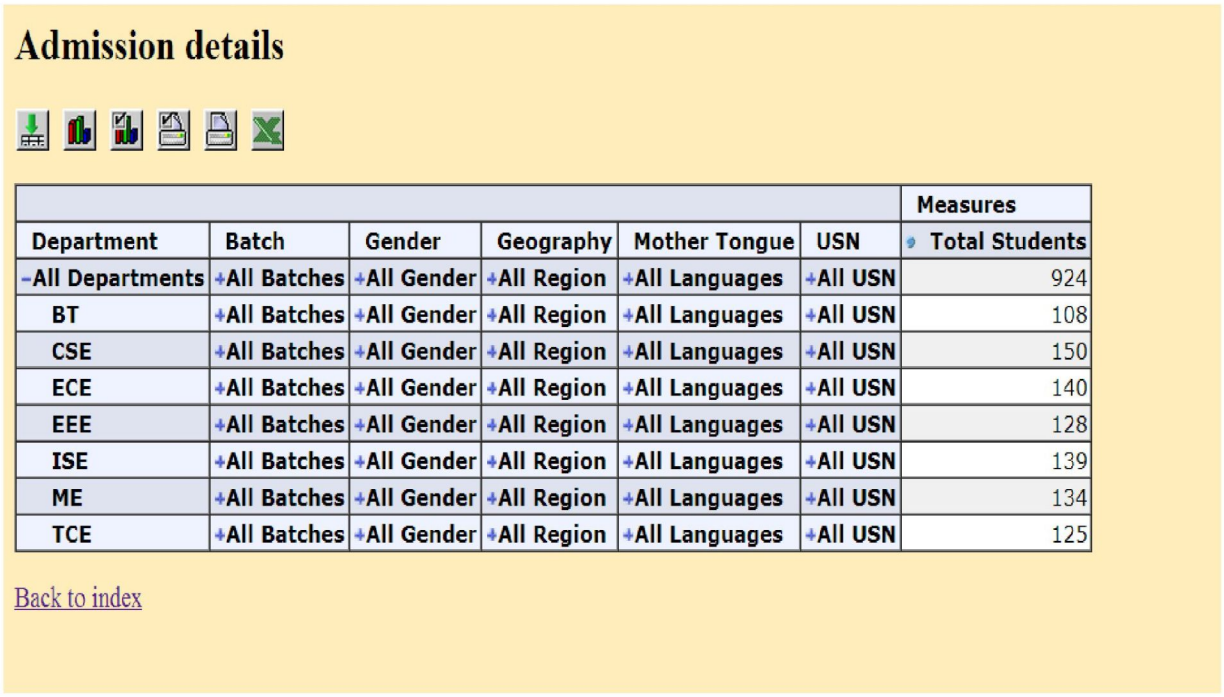


Figure 8.1: Homepage of SDW

The figure shows the homepage of SDW. It contains several options for the analysis of the student information. After selecting the "Admission details" option the user will be redirected to the page specified in figure 8.1.



Admission details

						Measures
Department	Batch	Gender	Geography	Mother Tongue	USN	Total Students
-All Departments	+All Batches	+All Gender	+All Region	+All Languages	+All USN	924
BT	+All Batches	+All Gender	+All Region	+All Languages	+All USN	108
CSE	+All Batches	+All Gender	+All Region	+All Languages	+All USN	150
ECE	+All Batches	+All Gender	+All Region	+All Languages	+All USN	140
EEE	+All Batches	+All Gender	+All Region	+All Languages	+All USN	128
ISE	+All Batches	+All Gender	+All Region	+All Languages	+All USN	139
ME	+All Batches	+All Gender	+All Region	+All Languages	+All USN	134
TCE	+All Batches	+All Gender	+All Region	+All Languages	+All USN	125

Back to index

Figure 8.2: Admission details page

The figure 8.2 shows the analysis part of the Admission details. When the user clicks on the drill-down ('+', just before the "All Departments"), the members of the "Department" dimensions will be displayed. Here "drill-down" is the accessing the members of the dimensions. In this case it shows different departments of PESIT will be displayed. The measure "Total Students" shows the total number of students admitted in the year of 2007. Here the aggregator "sum" is used to calculate the sum of all the students from different departments. The user can perform the drill-down operation on any of the dimensions mentioned.

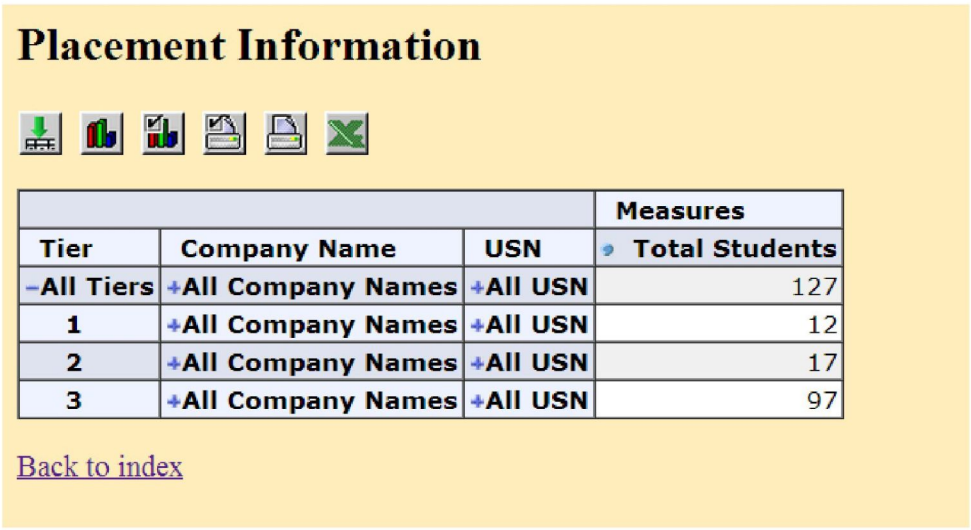
Placement Information

			Measures
Tier	Company Name	USN	Total Students
-All Tiers	+All Company Names	+All USN	127
1	+All Company Names	+All USN	12
2	-All Company Names	+All USN	17
	Allahabad Bank	+All USN	1
	Amartex	-All USN	3
		1PI07IS014	1
		1PI07IS015	1
		1PI07IS016	1
	Bajaj Auto	+All USN	1
	Balaji Telefilms	+All USN	1
	Bank of India	+All USN	2
	Bharat Aluminium Company	+All USN	1
	Bharat Electronics Limited	+All USN	3
	Bharat Petroleum	+All USN	1
	Bharat Sanchar Nigam Limited	+All USN	3
	Camlin Ltd.	+All USN	1
3	+All Company Names	+All USN	97

Back to index

Figure 8.3: Analysis of Placement Information

The figure 8.3 shows the placement statistics of the college for a particular department. It shows that there total of 17 students who got placed in a Tier-2 company. And "All Company Names" option on drill-down shows these tier-2 companies. Drill down option on tier-2 company shows the different tier-2 companies who visited for placements. And it also shows number of students placed for different companies. On clicking the drill-down option on "All USN" of company ", it shows the USN of the students who got placed in that particular company. In the above figure, it shows that 3 students got placed in the company "Amartex". On clicking the roll-up('-', just before the "All Company Names" for tier-2) option result in summary of this information. "roll-up" means going one level up. The result of roll-up operation is shown in the below figure.



Placement Information

			Measures
Tier	Company Name	USN	Total Students
-All Tiers	+All Company Names	+All USN	127
1	+All Company Names	+All USN	12
2	+All Company Names	+All USN	17
3	+All Company Names	+All USN	97

Back to index

Figure 8.4: Roll-up for placement information

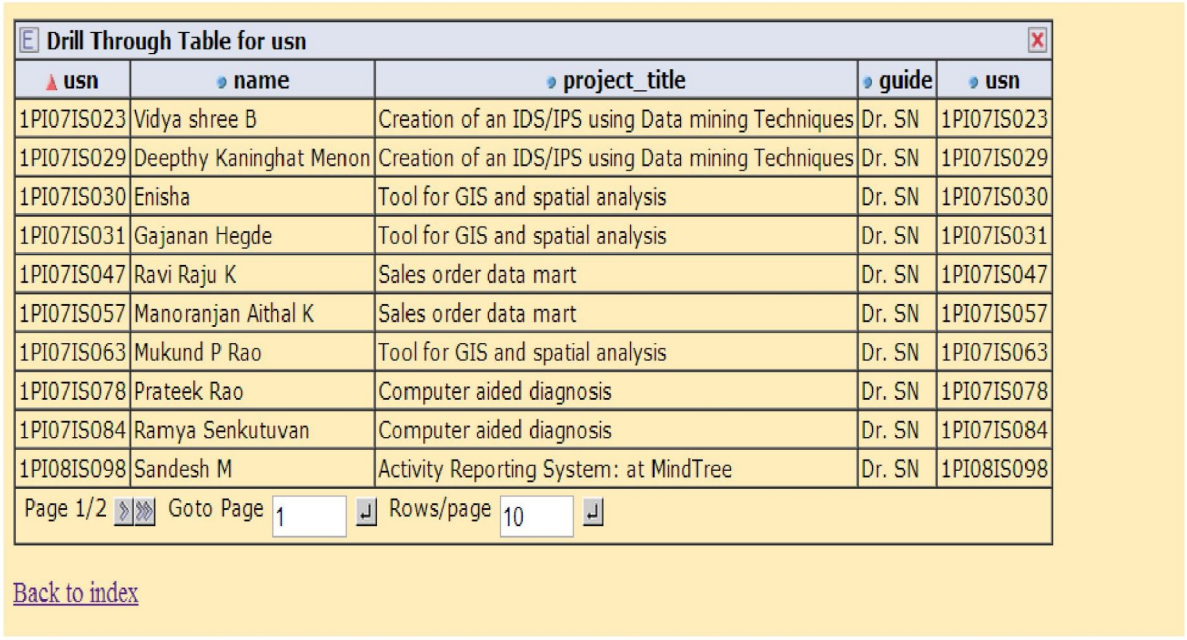
The figure 8.4 shows result of roll-up operation on "All Company Names" of tier-2 company. It only contains the abstract information like total number of students placed in tier-2 companies. The following figure is used to explain one of the OLAP operation called "drill-through".

Project Details

			Measures	
Guide	Project Title	USN	Number of Students	Grade
-All Guides	+All Projects	+All Students	138	8.971
AV	+All Projects	+All Students	9	8.778
CC	+All Projects	+All Students	10	8.2
COP	+All Projects	+All Students	9	9.111
CVP	+All Projects	+All Students	10	9.3
DHA	+All Projects	+All Students	3	9
DHJ	+All Projects	+All Students	8	9
Dr. SK	+All Projects	+All Students	6	9.667
Dr. SN	-All Projects	+All Students	12	9.583
	Activity Reporting System: at MindTree	+All Students	3	9.333
	Computer aided diagnosis	+All Students	2	10
	Creation of an IDS/IPS using Data mining Techniques	+All Students	2	10
	Sales order data mart	+All Students	2	10
	Tool for GIS and spatial analysis	+All Students	3	9
DVC	+All Projects	+All Students	5	9.2
GH	+All Projects	+All Students	3	8
JS	+All Projects	+All Students	11	9.182
KS	+All Projects	+All Students	10	8.9
MHR	+All Projects	+All Students	6	9.2
PR	+All Projects	+All Students	10	9
PRD	+All Projects	+All Students	2	8.5
Prof. SSS	+All Projects	+All Students	3	9.333
SHC	+All Projects	+All Students	3	8.667
SM	+All Projects	+All Students	10	8.1
UD	+All Projects	+All Students	2	9
VSP	+All Projects	+All Students	6	9.167

Figure 8.5: Project Details

The figure 8.5 shows the project details of department of information science. It has dimensions “Guide”, ”Project Title”, USN” and a measure called “Grade”. The result of drill-down operation on different dimensions are shown in the figure. The total number of students who have done project under a particular guide is calculated using the aggregator and it is mentioned under the measure “Number of Students”. The drill-through operation is used to browse the items. By clicking on any of the rows under the columns “Number of Students” or “Grade” will display the information about the students who have done a project under particular guide. Sample output of drill through operation is shown in the following figure.



Drill Through Table for usn

usn	name	project_title	guide	usn
1PI07IS023	Vidya shree B	Creation of an IDS/IPS using Data mining Techniques	Dr. SN	1PI07IS023
1PI07IS029	Deepthy Kaninghat Menon	Creation of an IDS/IPS using Data mining Techniques	Dr. SN	1PI07IS029
1PI07IS030	Enisha	Tool for GIS and spatial analysis	Dr. SN	1PI07IS030
1PI07IS031	Gajanan Hegde	Tool for GIS and spatial analysis	Dr. SN	1PI07IS031
1PI07IS047	Ravi Raju K	Sales order data mart	Dr. SN	1PI07IS047
1PI07IS057	Manoranjan Aithal K	Sales order data mart	Dr. SN	1PI07IS057
1PI07IS063	Mukund P Rao	Tool for GIS and spatial analysis	Dr. SN	1PI07IS063
1PI07IS078	Prateek Rao	Computer aided diagnosis	Dr. SN	1PI07IS078
1PI07IS084	Ramya Senkutuvan	Computer aided diagnosis	Dr. SN	1PI07IS084
1PI08IS098	Sandesh M	Activity Reporting System: at MindTree	Dr. SN	1PI08IS098

Page 1/2 Goto Page 1 Rows/page 10

Back to index

Figure 8.6: Drill Through operation

The figure 8.6 shows the USN, Name, Project Title and Guide for the selected "Number of Students".

The following figure shows example for decision support system (DSS) for the placement information used in the SDW.

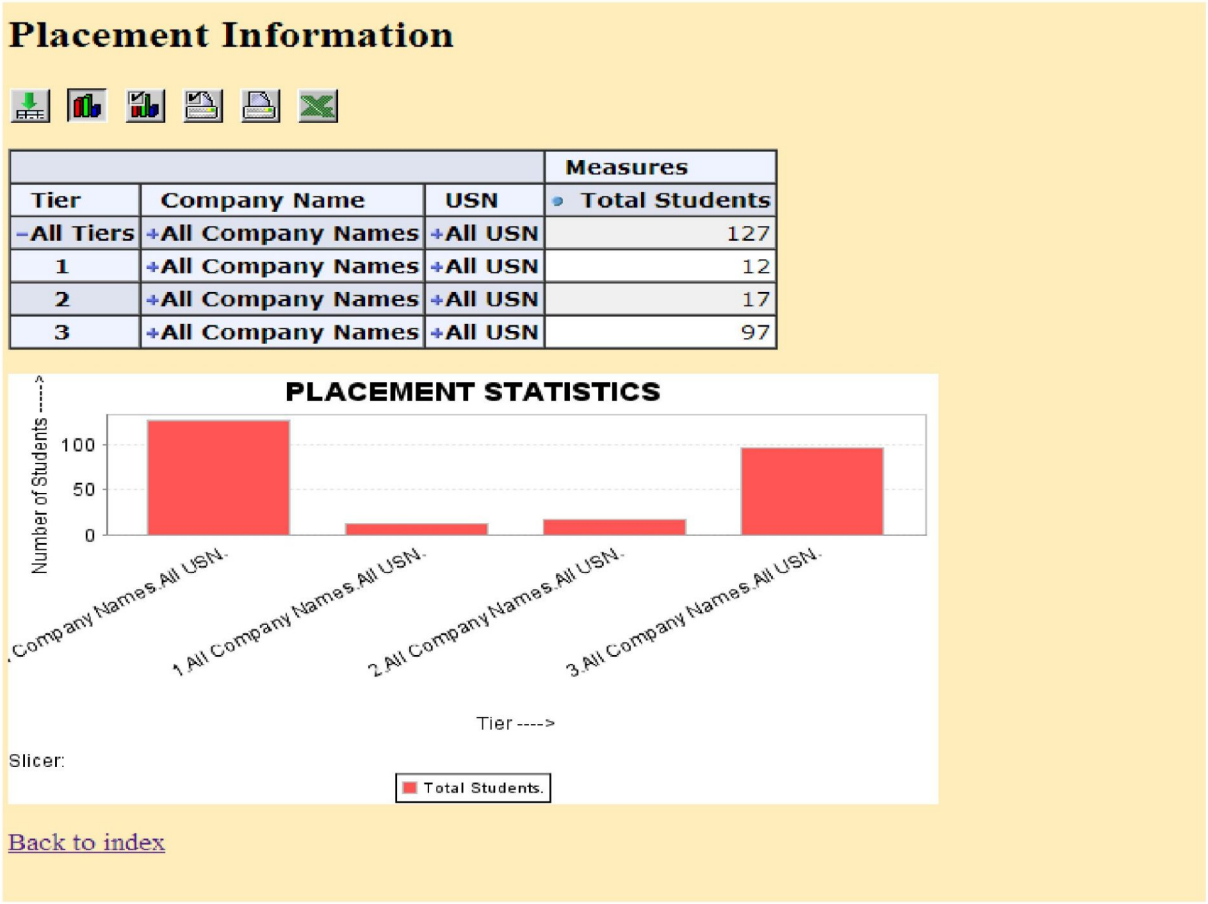


Figure 8.7: DSS for placement information

The figure 8.7 shows both analysis and decision support system for placement information. The X-axis represents different tiers and the Y-axis represents the number of students.