Relational Database Migration to Object Oriented Environment: A Re-engineering Approach

Tanveen Singh Bharaj ¹, Vishal Pandey ²

Student, Department of Computer Engineering, Jaypee Institute of Information Technology, Noida Sector - 128, Uttar Pradesh (U.P), India ¹

Student, Department of Information Technology, Maharaja Agrasen Institute of Technology, Rohini Sector - 22, New Delhi, India ²

ABSTRACT: The traditional relational databases (RDBs) have found applications in a number of areas and accepted as solution for storing and retrieving data. But RDBs cannot support data persistence, complex data structures and user defined data types required by object-oriented applications. On the other hand, object-oriented technology represents real world very well. Object-oriented databases (OODBs) aim to provide security, consistence, data independence and extensible data management services to support object-oriented models and extend support to handle complex and user defined data types which RDBs couldn't handle. It focuses on data rather than procedures and gives more security to data. OODBs support object oriented programming concepts of inheritance, polymorphism and encapsulation as well as database management concepts which are Atomicity, Consistency, Isolation and Durability (ACID). It is safe from unauthorized access as it provides three access specifiers which are private, public and protected which add strict security to data in database. The current study aims at designing an automated conversion tool that takes in relational database as input and converts it to object oriented schema by applying transformational rules and finally outputs an OODB. Then finally data migration takes place from RDBs to transformed OODB.

KEYWORDS: db4o, object oriented database, reengineering, schema transformation, data migration, object-oriented schema, relation databases.

I. INTRODUCTION

Huge volumes of data collected by organizations are often stored in relational databases systems (RDBs). But nowadays the volumes of data is increasing rapidly which can't be handled by RDBs. RDBs don't have capable storage structure to handle data such as digital, images, audio/video. There is absence of identity aspect and support towards user defined complex structures in RDBs [1]. They can't operate efficiently and effectively outside their language which is structured query language (SQL). Another drawback is that information is table form and relationships between entities are represented by values. In RDB exists relationship between data manipulation language (DML) and application programming language. Large scale commercial applications are developed in languages such as C++, java. In RDB, SQL is used as DML to insert, update and delete data. A major time gets invested in transforming data between two languages. This impedance mismatch adds complexities in RDBs [6]. In object-oriented databases (OODBs), programming language such as C++, Java itself acts as DML. Example, OODB may use java as DML and java itself can be used to build the complete application. Single language dependence helps in removing impedence mismatch and makes application development simple. As a response to overcome these shortcomings, and meet the increasing challenges in internet, OODBs have been proposed. Object-oriented database schema is considered to be conceptual description of the objects of real world. The object-oriented data model is considered to be the heart of the
coming generation of databases application [4, 11]. Users want to migrate their conventional databases by applying this new technology, without losing their data stored in legacy database systems. In OODBs data is represented as objects which are identified by object identifiers (OIDs). Objects reference each other by references/association. They allow object-oriented programmers to develop the product, store them as objects and replicate or modify existing objects to make new objects within the OODBs. There is requirement of OODBs in areas such as computer-aided design, software engineering and computer-integrated manufacturing [10, 13]. All these industries use OODBs to manage complex and highly interrelated data. Advanced office automation use OODB to handle hypermedia data. The efficiency of such database is greatly improved in areas where there is requirement of storing massive amount of data for example, a hospital database. The big O notation for such a database paradigm drops from O(n) to O(1) thereby increasing efficiency as one don't need to scan all tuples as opposed to RDBs. OODBs have many advantages over traditional database management systems as shown in Fig.1:

- Support reliable interaction with object-oriented languages such as Java and C++.
- Better handling of user defined and complex data types.
- Takes into consideration relationships that exists between the different objects.
- Programming can be done with small procedural differences without affecting the entire system.
- The operations defined on these kinds of system are not dependent on database application running at that moment.

Various OODBMS are - Db4o, Intersystem Cache, Objectivity/DB etc [8]. Among these, Db4o is an open source database which the current study focuses on. The current study aims to present a methodology on automating the process of creation OODBs incrementally from RDBs. The proposed approach provides semantic solution in schema transformation and data migration preserving the semantics and constraints of RDBs. The proposed methodology provides approach towards systematic migration which aims to meet the following needs:

1. Providing methodical support in schema migration of OODBs from relational one.
2. Assisting transformation process of relational schemas and databases.
3. Allowing users to extend support towards method for schema transformation.
4. Exploiting features of OODBs to the fullest.

II. RELATED WORK

The transformation of RDBs to OODBs has been a topic of study over the past few decades and several approaches have been proposed so far. Most of the earlier proposed approaches doesn't transform relational schema to object one automatically. Solution proposed in Pegasus [14] simply maps each relation to a class and relation instance as the class
extension. The drawback of this solution is that many object-oriented features are not exploited. Approaches proposed in [8] don't provide an automatic scheme transformation environment, the user is required to manually chose transformation rule to be applied on relational schema to outputs the corresponding object-oriented schema. This requires user needs to have a database and it's constraints in order to choose a correct rule to apply for transformation. Existing approaches do not yet regard data migration to OODB to the extent it is required to be. Solution proposed in [5] rather migrates database from relational to OODB by writing records/tuples into a file. The records are written in form of SQL query, this is a tedious task as records of database of every table has to be written manually and knowledge on SQL query is required. Error in manually writing record into file may make converted database migration and integration inconsistent. Schema transformation needs knowledge on relational schema semantics, such as
• Classification of relations and it's attributes
• Referential integrity constraints or
• Primary and possible candidate keys.
This information may not be directly fetched from relational schema, therefore semantic enrichment has to be implemented [8]. Semantic enrichment is seen as an important area of reverse engineering which is incorporated in the current study. The current research takes into consideration entire conversion process from RDB to OODB completely involving semantic enrichment, schema transformation, database integration and migration.
Study has been conducted to investigate the approaches and techniques required in conversion of database. In solution proposed by [2] there are three significant approaches for database conversion: objects view over RDBs, data integration and migration. The technique of translation can be widely classified into major two categories:
(i) Source-to-target technique: In this technique, physical source code is translated into an equivalent target with no Intermediate Conceptual Representation (ICR).
(ii) Source-to-conceptual-to-target technique: In this technique, the schema is interpreted from logical to a conceptual one by recovering the domain semantics and making them explicit.

III. ENVIRONMENT FOR MIGRATING TO OBJECT ORIENTED DATABASE SYSTEMS

General approach of conversion to OODB comprise of two step process of transformation. The first one, transforms the relational schema to object-oriented schema. And second one, migrates data from RDB to OODB [7]. During the process of conversion to OODB, it is necessary to maintain complex dependencies in data which maybe present in the system. Therefore, the current study proposes an approach which has an advantage over other comparable approaches that it extends it's support to both phases whereas other related approaches emphasize more on first phase. The major advantage of proposed solution is it's feasibility as object-oriented database model is a superset of the relational model [8]. The current research is divided into three steps process as illustrated pictorially in Fig. 3:
1. Schema Transformation and Extraction Phase : Maps relational schema to equivalent object oriented schema based on proposed transformation rules.
2. Instance Creation Process : Objects are instantiated from transformed object oriented schema and stored in created OODB.
3. Data Migration Phase : The transformed schema and objects instantiated over it gives support to migrate data to OODB from RDB.

These three phases are implemented using proposed automation tool that uses Db4o as the object-oriented database management system (OODBMS) to store converted RDBs and java as object oriented programming language to design the tool. The schema transformation phase, will reengineer the relational schema. The relational schema will undergo transformation to a well designed and understandable target schema being object oriented schema, to adapt to new applications. The input is relational database out of which relational schema will be fetched and outputs corresponding object oriented schema and it's OODB using proposed predefined set of transformation rules. Incrementally, transformation rule will be applied (based on input relational schema) that will transform a part of the relational schema, into a so called sub-pattern that will corresponds to a part of the object-oriented schema as shown in Fig. 2. The output
object-oriented schema will combine all tables together along with their attributes. Schema transformation and extraction rules are subdivided into class definition rule and attribute definition rule. The output of applying class definition rule is creation of java class structure from relations (tables). A corresponding object-oriented java file will be created for every relational table in RDB. And attribute definition rule will define the instance variables for that particular class. Further investigation in proposed approach will take into consideration primary and foreign keys, referential integrity constraints and additional dependencies while transforming to object oriented schema. This process of extracting additional information from relational schema is known as semantic enrichment [1]. The second phase which is instance creation process, will create and instantiate objects from their object oriented schema by accessing their corresponding java file created in first phase as illustrated in Fig 4.

The third phase is, database migration. It is a process in which schema translation and data conversion of source database (that is RDBs as per current study) takes place to an equivalent database in the target environment (that is OODB). Migrating to object technology consists of reverse engineering of the application programs and migration of the database [12]. As the current study proposes to create objects out of data, database migration provides greater flexibility in terms of reengineering input relational schema to corresponding object-oriented schema. Additional, higher levels of performance is achieved as data is not required to be converted at the runtime [1].

```
While ( No more rules can be applied)
{
	ransformation rule will be selected
transformation rule will be applied and
execute.
}
```

Fig. 2 Process of schema transformation

Data migration involves assigning values to the attributes of instances/objects created from instance creation process. These values are fetched from input RDB and migrated to objects created from second phase.

IV. THE MIGRATION ENVIRONMENT WITH AN EXAMPLE SCENARIO

The current study illustrates the proposed approach for conversion with the help of case study illustrated in Fig. 4. The proposed conversion algorithm is divided into three phases namely (i) Schema Transformation and Extraction phase (ii) Instance Creation process (iii) Data Migration phase.

Schema transformation is further divided into (a) instance definition rule and (b) attribute definition rule. The user will be queried to input it’s relational schema, specifying table/relation name along with all attributes for each table in the database. The source schema is translated to semantically enriched target schema. The transformation of a source schema inputted by user to a target schema consists of two subparts.
The first one, being Data-Base Reverse Engineering (DBRE), that aims to extract out the associated conceptual schema out of relational database, for example, Entity Relationship Model (ERM), expresses source schema’s explicit and implicit semantics. Relation/table, attributes, keys and dependencies in data form explicit semantics. It is very important to extract out additional semantic information that are not explicitly specified in RDBs. The second one, being Data-Base Forward Engineering (DBFE), that aims to extract out the resultant physical schema from the conceptual one obtained from the first part [2]. The proposed approach translate input schema to target one without any ICR.

IV. (a) Assumption and Dependencies
The automated tool where the proposed algorithm is implemented is designed in Java using db4o as OODB. Java has become one of the important programming language because of it's portability and popularity with internet applications [3]. For consistency purposes, the RDB input by the user for conversion should be in third normal form. If two relations have primary key whose names are same, the user is required to choose one table as parent table, from which other table is derived. In Fig. 5, underlined attributes refer that they are primary key to that relation.

IV. (b)Phase I: Schema Mapping Rules
These are transformation rules that defines how a particular construct/pattern in SQL will be mapped and transformed to it’s equivalent construct/pattern in OODB schema. The schema transformation is divided as:

Rule 1: Class Definition Rules
Every relation (table) of database is mapped to a new java class.

<table>
<thead>
<tr>
<th>Student</th>
<th>ID</th>
<th>name</th>
<th>dept_name</th>
<th>total_credits</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Takes</td>
<td>ID</td>
<td>course_ID</td>
<td>sec_ID</td>
<td>semester_year</td>
<td>grade</td>
</tr>
<tr>
<td>Section</td>
<td>course_ID</td>
<td>sec_ID</td>
<td>semester</td>
<td>year</td>
<td>room_no</td>
</tr>
<tr>
<td>Classroom</td>
<td>building</td>
<td>room_no</td>
<td>capacity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course</td>
<td>course_ID</td>
<td>title</td>
<td>dept_name</td>
<td>credits</td>
<td></td>
</tr>
<tr>
<td>Department</td>
<td>dept_ID</td>
<td>dept_name</td>
<td>HOD</td>
<td>building</td>
<td></td>
</tr>
<tr>
<td>Instructor</td>
<td>instr_ID</td>
<td>name</td>
<td>dept_ID</td>
<td>salary</td>
<td></td>
</tr>
<tr>
<td>Teaches</td>
<td>instr_ID</td>
<td>course_ID</td>
<td>sec_ID</td>
<td>semester</td>
<td>year</td>
</tr>
<tr>
<td>Exchange_Student</td>
<td>ID</td>
<td>Country</td>
<td>from</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5 University Database Schem

Pattern:

<table>
<thead>
<tr>
<th>Student</th>
<th>ID</th>
<th>name</th>
<th>dept_name</th>
<th>total_credits</th>
</tr>
</thead>
</table>

Fig. 5(a) Student table
Referring to Fig. 5(a), student table will be mapped to Student class as given below:

Result:
Class Student
{
}

**Rule 2:** Class Inheritance Rule

If two tables in a database have the same key (primary key), then class inheritance rule can be applied where one table will be chosen as the parent table and the other as a child by the user. Here user intervention is required to choose the parent table out of two tables with the same keys.

**Pattern:**

<table>
<thead>
<tr>
<th>Relation</th>
<th>ID</th>
<th>name</th>
<th>dept_name</th>
<th>total_credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exchange_Student</td>
<td>ID</td>
<td>Country_from</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5(b) Student and Exchange_Student table

Referring to Fig. 5(b), if user chooses Student as the parent table then the result would be:

Result:
Class Student
{
}
Class Exchange_Student extends Student
{
}

**Rule 3:** Attribute Definition Rule

Attributes of relation are mapped as instance variables of that class (derived from rule 1 or rule 2).

**Pattern:**

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Datatype</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>INT</td>
</tr>
<tr>
<td>Name</td>
<td>VARCHAR</td>
</tr>
<tr>
<td>Dept_name</td>
<td>VARCHAR</td>
</tr>
<tr>
<td>Total_credit</td>
<td>INT</td>
</tr>
</tbody>
</table>

Fig. 5(c) Student table

Referring to Fig. 5(c), student table will be mapped to Student class as given below:

Result:
Class Student
{
    int ID;
    String name;
    String dept_name;
    int total_credit;
}

**Rule 4:** Mapping foreign key relationship

Every foreign key relationship in relational schema is mapped to object identifier references or association attributes. If one table has a key referencing primary key of another class, then in OODB, the child class will have object of parent class as its instance variable to keep referential integrity constraint intact. The foreign key attribute is dropped from the class, leaving the class with semantically meaningful attributes and association attributes with other classes. The foreign key mapping rules have been described in rule 5 and rule 6.
Rule 5: Mapping many-many relationship
A relation whose attributes are a concatenation of primary keys of two other relations form many-to-many relationships or binary many-to-many relationship. If there is a table that is formed by combination of primary keys of two separate relation, the class thus formed by mapping, of such tables will contain reference to other two classes.

Pattern:

<table>
<thead>
<tr>
<th>Student</th>
<th>ID</th>
<th>name</th>
<th>dept_name</th>
<th>total_credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Takes</td>
<td>ID</td>
<td>course_ID</td>
<td>sec_ID</td>
<td>semester</td>
</tr>
<tr>
<td>Course</td>
<td>course_ID</td>
<td>title</td>
<td>dept_name</td>
<td>credits</td>
</tr>
</tbody>
</table>

Fig. 5(d) Student, Takes and Course table

Referring to Fig. 5(d), Student, Course and Takes table will be mapped as given below:

Result:

Class Student
```
{  
  int ID;  
  String name;  
  String dept_name;  
  int total_credits;  
}
```

Class Course
```
{  
  int course_ID;  
  String title;  
  String dept_name;  
  int credits;  
}
```

Class Takes
```
{  
  int sec_ID;  
  int semester;  
  int year;  
  String grade;  
  set (Student S);  
  set (Course C);  
}
```

While setting objects of Student class and Course class during initialization of Takes class object, only those ID and course_ID will be accepted for insertion, which have already been set and inserted in previously initialized objects of Student and Course class.

Rule 6: Mapping one-one/ one-many relationship
A dependency between the keys of two table/relation in both directions leads to a one-to-one relationship. This is implemented by adding object identifier references of each other in both derived class.
Referring to Fig. 5(e), Department and Instructor table will be mapped class as given below, where dept_ID in Instructor table is foreign key to Department table:

**Result:**

```java
class Department {
    int dept_ID;
    String dept_name;
    String HOD;
    String building;
    Instructor inst;
}

class Instructor {
    int Instr_ID;
    String name;
    int salary;
    Department dept;
}
```

Since every instructor belongs to a department and a department has set of instructors, therefore we need to set object reference of both classes in each other. During initialization of Department class object, while setting object of Instructor inside Department class object only those instr_ID will be accepted that have been previously assigned in objects of Instructor class. Similarly during initialization of Instructor class object, while setting object of Department class inside Instructor class object only those dept_ID will be accepted that have been previously assigned in objects of Department class.

**Rule 7:** Mapping composite multivalued attribute

Composite multivalued attribute tuple/row of atomic values will be transformed into a set of instance variables containing the composite attributes. Here user intervention and interaction is essential to specify datatype and name of composite attribute derived from multivalued attribute.

```java
class Student {
    int ID;
    String name;
    String dept_name;
    int total_credits;
    set address as {
        String street;
        String city;
    }
}
```

Referring to Fig. 5(f), student table will be mapped to Student class as given below with multivalued attribute:

**Result:**

```java
class Student {
    int ID;
    String name;
    String dept_name;
    int total_credits;
    set address as {
        String street;
        String city;
    }
}
```
While transforming Student relation to Student class, user shall be prompted about any multivalued attribute. If it exists, the user shall be required to specify which one is multivalued attribute and specify the subdivided attributes of multivalued attribute in transformed class.

E.g. In Fig. 5(f) user will be required to specify subdivided attributes of Address as String street and String city.

IV. (b) Phase II: Instance Creation Process
Every Java class file created from schema mapping rules will be compiled containing the structure of class along with the default and parameterized constructors. The files will be compiled at run time using reflection library in Java. This protects data from illegal access as only class objects can access other members of that class and only through constructors one can access and assign data values. Run time compilation of Java class is required for persisting objects of classes defined in java files. These objects are the one stored in created OODB in Db4o. This follows data migration into created OODB from RDB.

IV. (c) Phase III: Data Migration Phase
Data stored in tuples of RDB are transformed into complex objects in OODB. This involves loading relational data to the automated tool, and then inserting it into output database produced as a result of schema transformation and instance creation process. Once the object-oriented schema is ready and corresponding java files have been compiled. The user is prompted to transfer data from relational to OODB created in db4o. The user will specify the table for which he wants to migrate data, the tool will automatically generate required text fields that are equal to number of instance variables in that class. If the tool reads that input class has object reference to another class, it will automatically generate textfields for that reference class too for the user to input records. While entering the records the tool will take into consideration, no database constraint is violated. Example (considering foreign key constraint): If a class has object reference to another class, then reference object can take values that exists in the OODB for that reference class as illustrated in Fig.6, Fig. 7(a) and Fig. 7(b).

In Fig. 6. Structure of Car and Pilot class are as follows:

```java
class Car {
    String model;
    Pilot pilot;
}
class Pilot {
    String name;
    int points;
}
```

![Fig. 6](image_url) Insertion of values in Car table

After insertion of values in Car class, Fig. 7(a) shows structure of Car table in OODB. Car table has reference to pilot class. Fig. 7(b) shows structure of each record inside Car table.
Fig. 7(a) Car table values in OODB

Fig. 7(b) Structure of Car table

Fig. 8 shows that values inside transformed Pilot class in OODB. From the figure it is observed that the record, entered inside Car class in Fig. 6 already had a reference in Pilot class table in Fig. 7(b). Had the record being entered in Car class inside Pilot reference was missing in Pilot table error would be generated as shown in Fig. 9. Fig. 9 generates error as reference to Pilot("Tanveen",101) is missing in Pilot table in Fig. 8. Therefore proposed data migration process not only migrates data from relational to transformed OODB, but also ensures same consistency is achieved as it was in RDB.

V. THE ARCHITECTURE OF THE CONVERSION TOOL

The users will be provided with tool support that implements the current study. During conversion of database transformation rules are implemented depending on RDB patterns. Some migration rules require user interaction, but the current study have tried to minimize it to maximum. Therefore the user has to support conversion tool to simplify the migration process and vice-versa. The proposed conversion tool implements the following tasks:

1. Identify which rule to apply for a relational pattern
2. Applying transformation rule after identification.
3. Initialising objects for transformed pattern.
4. Visualizing the migration process.
5. Applying the migration process

These different tasks are reflected by the architecture sketched in Fig. 10. The lowest level is called Transformation Definition where the transformation rules have been defined. The function of the next level, Knowledge Application, is that it determines according to given relational pattern which transformation rule should be chosen and applied. After application of rules on entire input relational schema incrementally java file of every relation will be created and compiled. Certain rules defined in this study also require user interaction which will be identified and prompted to user for his response in this level.
Migration Process Control instructs the user and records the progressing process of migration. Therefore user gets an outline of the migration technique, that is which section of the schema have been transformed or still have to be transformed. Furthermore, it is enables the user to undo certain portions of migration process. Finally, the level Object Representation visualizes all items in newly created ODB in db4o after application of required transformation rules and completion of migration process.

VI. CONCLUSIONS AND RESULTS

The current study proposed an interactive environment for transforming relational into object-oriented database schemas with automated method for data migration to converted OODB. The user will input the source relational schema, depending on input relational pattern transformational rules are applied to get target enriched object-oriented schema. The transformation rules work on pattern to be fetched and matched from relational schema, a set of conditions and a sequence of operations to take place for transformation [8]. Such conversion can map database structure and it's semantics from RDB to OODB without any information loss. The referential integrity constraints such as primary key and foreign key constraints have been rightly incorporated in object schema. The transformed schema is java class, which is compiled at run time to get .class file. Objects are initialized from .class file to construct OODB for the input RDB in db4o. Finally data is migrated to the transformed OODB. The current research is useful in incorporating semantics like generalization, association in transformed schema. Fig. 11 represents the comparative study between RDB and obtained OODB from RDB, thus the need to migrate to OODB from RDB. The contribution of current research is that:

1. Providing model for the adaptable definition of transformation rules,
2. Proposing transformation rules that utilise most of the object-oriented features
3. Ability to produce database migration programs from transformed object oriented schema.
Fig. 10 Architecture of Proposed Conversion tool

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Relational Database</th>
<th>Created Object-Oriented Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support towards user datatypes</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Joins</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Handling long transactions</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Inheritance, Polymorphism &amp; Encapsulation</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>ACID (Atomicity, Consistency, Isolation &amp; Durability)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Impedance Mismatch</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Speed</td>
<td>Slower</td>
<td>Faster</td>
</tr>
</tbody>
</table>

Fig. 11. Comparative study between RDB and transformed OODB

REFERENCES


