A Comparative Study of Tuple Timestamped Data Models

Dr. Shailender Kumar  
Department of Computer Science and Engineering  
Ambedkar Institute of Advanced Communication Technologies & Research, Geeta Colony, Delhi, India

Disha Aggarwal  
Department of Computer Science and Engineering  
Ambedkar Institute of Advanced Communication Technologies & Research, Geeta Colony, Delhi, India

Abstract: Time is one of the most challenging facets to handle in real world applications. Most applications access or manipulate temporal data so such applications rely on temporal databases which record the time varying data. The objective of this paper is to provide a concise overview on concepts of temporal database and review the work in the field of tuple timestamping data models. The paper introduces the basic of conventional database and then compares it with the temporal database. The paper evaluates different time dimensions and query languages along with them. A comparative analysis of different temporal models is done based on some parameters. Finally, all the temporal models are summarized and a conclusion is given with the future work in the field of temporal database.

Keywords: database; temporal database; attribute timestamping; tuple timestamping; transaction time; valid time; data models

I. INTRODUCTION

The current interest in the relational approach to database is because of the Codd, who presented the relational model which includes the relational structure and its associated advantages. But this relational model does not consider the temporal dimension as it does not keep a track of past or future aspects of time in database but only focuses on the bivalent Boolean logic. Applications like banking, inventory management, healthcare management, reservation systems, insurance applications, weather monitoring, etc. involves the data that changes with time and it is important to keep record of all that information.

A database which stores different time aspects (past, present and/or future) of the data is known as temporal database. In this section we discuss the basic database concepts and how the conventional databases are different from the temporal databases.

A. Basic Conventional Database Concepts

A database management system (DBMS) is a collection of program that enables users to create and maintain a database. A database is a collection of related data. By data, we mean known facts that can be recorded and that have an implicit meaning [2]. So the main objective of DBMS is to provide a convenient and effective method of defining, storing, retrieving and manipulating the data contained in the database.

Data can be divided into: Past Data, Present data and Future Data. The data about the past events and circumstances which existed before in the database is referred as Past Data.

The data that is valid for current time is known as Present Data and sometimes denoted with “now” keyword. The data which is generated over a specific time range, let say, a day or week ahead of the current time is known as Future Data.

To structure this data, data models are needed. Data Model can be defined as an integrated collection of concepts for describing and manipulating data, relationships between data, and constraints on the data in an organization. Data Model includes two types of schema: Physical Schema and Logical Schema. Physical Schema describes the database design at the physical level, while the Logical Schema describes the database design at the logical level [2].

These data model schemas must exhibit data independence. Data Independence is the ability to change the schema at one level of the database system without having to change the schema at the other levels [11]. Data Independence is of two types: Physical Data Independence and Logical Data Independence. Physical Data Independence is the ability to change the internal schema without affecting the conceptual or external schema. Logical Data Independence is the ability to change the conceptual schema without affecting the external schemas or application programs.

Conventional Database Data Models can be classified as:

i. Hierarchical Database Model: This data model organizes the data in a tree-like structure in which each child node can have one parent node only and is connected to one another through links.

ii. Relational Model: This data model maintains data in the form of tables, also known as relations, consisting of rows and columns.

iii. Network Model: This data model organizes the data in the form of graphs and all the nodes are linked to each other, without any hierarchy, using links.

iv. Entity-Relationship Model: This data model is composed of entity types (objects) and specifies relationships that can exist among those objects.
of different data models. Section 4 states the conclusion with the future work.

II. TEMPORAL DATABASE CONCEPTS

A. Time Dimensions

Temporal database supports three types of time dimensions, which exhibits independence from each other: user-defined, valid and transaction time. User-defined time is a time representation designed to meet the specific needs of users.

Valid time specifies when certain conditions in the real world are, were or will be valid. Valid time can be represented with single-chronon identifiers (event timestamps), with intervals (as interval timestamps), or as valid-time elements, which are finite set of intervals [4].

Transaction time automatically captures changes made to the state of time-variant data in a database. This time dimension represents the time period during which an instance is recorded in the database. Transaction time is mostly used to support versioning, which generally implies an object-oriented data model [14].

There is also another form of time-variant data, called bitemporal data. Bitemporal data are union of valid time and transaction time data. Several data models concerning bitemporal data are known in the literature: The Bitemporal Conceptual Data Model (BCDM) is a very simple model capturing the essential semantics of time-variant relations. Another example of a bitemporal model is Nested Bitemporal Relation Data Model (NBRDM). The model, as the name implies, is based upon a nested relational schema.

B. Timestamping

The temporal relational data models are categorized according to the timestamps attached and the time dimension they support. The data that have time-related information are known as temporal data.

A timestamp is a time value associated with a data value that might be a temporal element, time interval, or time point [18].

There are two types of timestamping:

1. Attribute Timestamping:

In this type of timestamping, value of timestamp is attached with the attributes of the database table. The various characteristics of attribute timestamping:

i. It uses non-first normal form to store time-variant data.
ii. It avoids redundancy and is more expressive because it doesn’t contain the repetition of the multiple rows in the table due to change in the attribute values over the time.
iii. It uses single relation/tuple to store time varying attributes. It doesn’t splits the time-variant and time-invariant data into several tuples.
iv. It may not be capable of efficiently using existing storage structure. Hence, it cannot use high effective relational techniques, such as Query Evaluation.
v. It performs better for queries with low levels of nesting because as the nesting increases, it will be more
complex to apply timestamp values with those attributes that are divided further in more sub-attributes.

vi. Attribute Timestamping is more suitable to execute valid time queries as it executes them faster.

vii. This timestamping is least suitable for bitemporal queries as it executes them slowly.

### EMPLOYEE TABLE

<table>
<thead>
<tr>
<th>EmpID</th>
<th>Ename</th>
<th>Sal</th>
<th>T Value</th>
<th>T Start</th>
<th>T End</th>
</tr>
</thead>
<tbody>
<tr>
<td>E123</td>
<td>Rajiv</td>
<td>30000</td>
<td>12/02/2012</td>
<td>15/02/2014</td>
<td>20/02/2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35000</td>
<td>15/02/2014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E234</td>
<td>Simran</td>
<td>35000</td>
<td>15/03/2013</td>
<td>20/03/2015</td>
<td>25/03/2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>40000</td>
<td>20/03/2015</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3. Attribute Timestamping**

**2. Tuple Timestamping:**

In this type of timestamping, value of timestamp is attached with the tuples of the database table. The various characteristics of tuple timestamping:

i. It is based upon the first normal form (1NF) to store time-variant data.

ii. It contains high data redundancy because of the repetition of multiple rows of a table due to change in attribute values over the time.

iii. It uses relational tables to represent time-invariant as well as time-variant data. It splits both time-variant and time-invariant data into several tuples.

iv. It is capable of efficiently using existing storage structure. Hence, it can use high effective relational techniques.

v. It performs better for more complex nesting structure of the query as it adds the timestamp value to the row of the table. So, it results in less complex structure after adding timestamp value.

vi. Tuple Timestamping is less suitable to execute valid time queries as it executes them slowly.

vii. This timestamping is more suitable for bitemporal queries as it executes them three times faster than attribute timestamping does.

### EMPLOYEE TABLE

<table>
<thead>
<tr>
<th>EmpID</th>
<th>Cou nt</th>
<th>Ename</th>
<th>Sal</th>
<th>T Start</th>
<th>T End</th>
</tr>
</thead>
<tbody>
<tr>
<td>E123</td>
<td>0</td>
<td>Rajiv</td>
<td>30000</td>
<td>12/02/2012</td>
<td>15/02/2014</td>
</tr>
<tr>
<td>E123</td>
<td>1</td>
<td>Rajiv</td>
<td>35000</td>
<td>15/02/2014</td>
<td>20/02/2016</td>
</tr>
<tr>
<td>E234</td>
<td>0</td>
<td>Simran</td>
<td>35000</td>
<td>15/03/2013</td>
<td>20/03/2015</td>
</tr>
<tr>
<td>E234</td>
<td>1</td>
<td>Simran</td>
<td>40000</td>
<td>20/03/2015</td>
<td>25/03/2016</td>
</tr>
</tbody>
</table>

**Figure 4. Tuple Timestamping**

### III. TEMPORAL DATA MODELS

#### A. Introduction

The temporal models are used to manage both types of data: time-variant as well as non-time variant. Time can be added to any of the data model like entity-relationship model, semantic data models, knowledge based data models and deductive databases. But most of the work has been done on relational data models and object-oriented data models only. That’s why in this paper, temporal aspects of relational and object-oriented data models are discussed only.

In conventional database systems, the management of time-variant data is done entirely at the level of user-defined time in which the attribute values are drawn from a temporal domain. The SQL-92 Date, which is a tuple calculus-based language, has granularity of a day; a Time has a granularity of a second, having a range of 100 hours only. A Timestamp combines the range of the Date with the granularity of a second (there are Timestamp variants with a granularity of fractions of a second) [3].

1. **Valid Time:** Temporal Data Models can be compared on the basis of the valid time dimension. This can be done as there are three ways to represent the valid time in any temporal data model. These are: single chronon (event timestamp), intervals (interval timestamp), or as valid-time elements (finite set of intervals) [4]. In temporal data models the valid time can be associated in three ways i.e. with the values of individual attributes, attribute groups, or with the entire tuple [1]. Some of the examples of valid time models are: Ariav, HDM, Lum, Sadeghi, Jones, Navathe, Snodgrass, Sarda, and Yau.

2. **Transaction Time:** Object oriented data models are generally identified by implementing versioning on it and transaction time supports the versioning [1]. These data models allow arbitrary, user-defined identifiers to be associated with versions. An entire version hierarchy can also be associated with a version if required by a data model. Some of the examples of transaction time models are: BCDM, Ben-Zvi, Lomet, ADM, Snodgrass, and Postgres.

3. **Set of Temporal Data Models:** A temporal data model should meet the certain goals, which can lead to the best possible outcomes. The application which is going to be modelled should have its semantics captured clearly and precisely. It should extend the already existing data model, like relational model, so that it will produce more consistent and accurate outcomes. It must represent all the time-variant and time-invariant behaviour of the objects logically. But, this is probably not possible to have all these features while designing a temporal data model.

There exists an extension of SQL i.e. TSQL2 which uses Bitemporal Conceptual Data Model as its underlying data model which is used to capture the essential semantics of time-varying relations. A different model i.e. representational data model is utilized for implementation with assured high
performance. Another data model i.e. presentational data model depicts the time-varying behaviour to the user. So, no single data model can achieve all the goals that is why a set of data models should be used in order to fulfill all the desired requirements with best possible outcomes.

B. Query Languages
A data model also contains a set of operations on objects, in combination with set of objects having structures and a set of constraints on those objects. This set of operations is known as the temporal query language which is used to access, modify, and record the objects in data model [10]. Some of the query languages are: ILs, an intensional logic devised in computational linguistics [3]; Quel, defined for Ingres relational DBMS, is a tuple calculus-based query language; relational algebra is a procedural language, denoting relations as objects; SQL, a language of conventional relational database, is a tuple calculus-based language.

1. User-Defined Time: This time aspect supports both the data models and query languages. Most of the commercial relational DBMSs supports user-defined time. Time can be treated as an abstract data type with the help of its own predefined set of operations. Postgres, being an object-oriented query language, supports this approach.

2. Valid Time: As valid time supports the future aspects of certain data besides the past values of that data, it has greater influence than the user-defined time. Valid-time can be added to the data models and on the query languages with few different approaches. The foremost approach is to directly utilize the extensively responsive ability of relational or object-oriented data model. This approach is the easiest one as it will not require any changes to the existing data model or the query languages. But, there occurs one problem in this approach that it makes the query optimization extremely more difficult. This happens because of the lack of suggestions provided by the query language, whether the access methods that are concerned with time-varying values should be employed or not.

In the next approach, data models and query languages can be extended with the required needs by showing their support to the time-varying figures. This approach cannot be used for relational data query languages. The problem of query optimization is still present in this approach as user-defined functions do all the manipulations related to time values. The former approach is used widely by temporal relational query languages. According to this, the usage of distinctive data model concept and query languages can support the information related to variation of the valid time to the great extent.

3. Transaction Time: Transaction Time defines the state of time when the fact is logically entered in the database. It does not indicate the validity of the fact in either respect of past or future time. It is used when one wants to roll back to the time when a particular fact was stored in the database irrespective of its validity. Transaction time supports the following type of versioning:

i. Extensive Versioning
In this type, versioning of tuples, object instances or attributes is done by themselves. In support of this versioning, there are different approaches same as with the valid time. In the first one, data models are used directly without doing any changes in them or in the query languages. Transaction time can be adapted irrespective of the time semantics in the database. The next approach extends the data model and query language so that the information related to time-varying data can be supported. In the last approach, some modifications are done in the existing data model and query language according to the required needs so that it can support time-varying data. Transaction time can be represented as a graph and not as a tree because it consists of branching of time whenever the objects are versioned, in addition to the feature that new version is created by merging the two old versions.

ii. Schema Versioning
A schema of an application can be modified according to the variation in the application’s needs. So, multiple schemas can be altered logically in schema versioning. It supports the versioning of the definition of the objects in the database. Schema versioning can be applied in both relational databases and object-oriented databases.

The essential point that needs to be kept in mind is that if extensive versioning is applied then support for schema versioning may or may not be present. But, if extensive versioning is not adopted then there exists no relevance in using schema versioning there.

C. Data Models
In temporal data models, there exists some relational data models and object-oriented data models. A brief study is done on the previous data models, which get out-dated with time, just to get an overview of the work that has been done related to temporal data models so far.

i. Ariav proposed a data model named as Temporally Oriented Data Model [13]. The researcher uses relational data model as the base model, incorporating both the time dimensions that is, valid time and transaction time. This model uses the single chronon (event timestamp) representation of time. It supports the storage and retrieval of data through the use of a query language, SQL.

ii. Ben-Zvi proposed a Time Relational Model (TRM) in order to model the different aspects of time [15]. TRM uses first normal form relations and supports both type of time dimensions (valid and transaction time). This model uses intervals, which are pairs of chronons, to represent time in the relational data model used.
iii. There exists a model named as DATA, which is based on relational data model. This model uses transaction time dimension only to store the time at which data was entered in the database. This model uses a single chronon representation of time to timestamp the value. It is only a proposed model which has no implementation.

iv. Another model is DM/T which uses relational data model as a base model. Timestamping is done using one dimension only that is, transaction time. This model does not deal with the future aspects of time. Time representation is done by applying event timestamps in database. Relational Algebra is used as a language to retrieve the data from database.

v. There is another model, Historical Data Model, which stores the history of data so that user can use the previous data by rolling back it to the specific period of time. This is done using the relational data model having valid time as the time dimension to store the validity of data also. Time is represented as single chronon or event timestamp and uses a query language, ILs (intensional logic).

vi. Postgres is another model which is based on object-oriented data model. This model uses transaction time dimension to store the previous data but does not stores the future aspect of time. Time is represented in the form of intervals (pair of chronons) and uses Postquel language to retrieve and store the data, which is based on a language Quel (tuple-calculus based query language).

vii. Navathe proposed a data model named as Temporal Relational Model, based on the relational data model. This model can be expressed as temporal extension of the relational data model. This model uses valid time dimension so as to store validity of the data in near future and the representation of this valid time is done using the intervals of time. This model uses a temporal extension of SQL known as TSQL.

viii. Segev presented implementation of Temporal Data Model which is based on the relational data model [17]. It is implemented using valid time dimensions and represented as single chronon. To access the data, this model uses TDM language which is based on SQL.

ix. Weiderhold proposed a data model, Time Oriented Databank Model, having relational data model as underlying model. This model is based on valid time dimension with the representation of time as single chronon.

x. C.S. Jensen and R.T. Snodgrass explained a temporal approach named as Bitemporal Conceptual Model (BCDM) [6]. It supports both types of time dimension (valid and transaction time) and is linear and discrete in nature. The main advantage of this model is that it remains simple and understandable while storing the temporal data. But, it is not suitable to manage and implement large number of tuples.

xi. Debabrata Dey, Terence M. Barron and Veda C. Storey suggested Temporal Event Entity Relationship Model (TEERM) in 1995 [7][8]. It is bitemporal in nature. This model captures the real world aspects and introduced new relationships like static relationship, dynamic relationship and quasi-static relationship. But, representation of events can sometimes introduce redundancy into tables.

xii. Nina Edelweiss, Patricia N. Hubler, Mirella M. Moro and Giovani Demartini represented implementation of Temporal Functionality in Objects with Roles (TF-ORM) data model [5]. It is an object-oriented data model but differs from other object-oriented data model in a way that it uses a role concept to represent the different behaviours of an object. There can be several roles present for each class, and each role will depict a different behaviour for the object of that class. It uses TF-ORM query language as its language to retrieve and access the data.

xiii. Jan Mate and Jiri Safarik presented an implementation which uses an automatic rule based schema transformation technique to convert relational database to temporal database [9]. It uses schema versioning and versioned table techniques. Implementation of this model is done on PostgreSQL database with SQL as query language and PL/SQL as implementation tool language.

xiv. Vincent S. Lai, Jean-Pierre Kuilboer and Jan L. Guynes suggested a model as an extension of Enhanced Entity Relationship (EER) [12][8]. This model is named as Temporal Enhanced Entity Relationship (TEER), which uses both time dimensions (valid and transaction time) and stores the full history of every entity and its associated relationships. This model can easily be mapped to an extension relational model.

xv. Gadia-3 identified a temporal extension to SQL model, which was proposed on the basis of convention database model given by Navathe [4] [16]. It is a valid time model which allows both the attributes that is, time-varying and non-time varying but they must all be of the same type. It can cope with the multiple occurrences of value equivalent events in the same partition of the timeline.

**D. Comparative Analysis**

The tuple timestamped data models are compared as per the following parameters:

i. Time Dimension: This characteristic defines the different time aspects of models whether they use valid time, transaction time or both of them.

ii. Implementation of Model: This characteristic defines that whether the model has implementation or not.

iii. New Proposed Language: This depicts that on which convention database that temporal model is proposed.

iv. Implementation Language: This specifies a language that was used to implement that model.

v. Extended Non-Temporal Model: This criteria will depict that on which convention database that temporal model is proposed.

vi. Language used to Query: This depicts a language that is used to access and manipulate the data in the particular model.

vii. Techniques used to manage the Data: This field enumerates the various data management techniques used to manage the data in the models.

viii. Advantages: This specifies the various advantages that the model persists.

ix. Disadvantages: This specifies the different disadvantages that the model has.
A comparison of some prominent tuple timestamped temporal data models is done as shown in the Table I.

<table>
<thead>
<tr>
<th>Name of the model</th>
<th>Time Dimension</th>
<th>Implementation Of the Model</th>
<th>New Proposed Language</th>
<th>Implementation Language</th>
<th>Extended Non-temporal model</th>
<th>Language Used to Query</th>
<th>Techniques used to manage the data</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCDM</td>
<td>Bitemporal</td>
<td>×</td>
<td>×</td>
<td>Relational Model</td>
<td>×</td>
<td>tf orm</td>
<td>(a) Easy to understand (b)Flexible representation (c)It's graphical representation is easy to understand (d)Simple to understand and expressive in nature</td>
<td>(a)Easy to understand (b)Stores full history in one tuple</td>
<td>Difficult to handle large number of tuples</td>
</tr>
<tr>
<td>TEERM</td>
<td>Bitemporal</td>
<td>×</td>
<td>×</td>
<td>Extended ER model</td>
<td>×</td>
<td>tf orm</td>
<td>(a) Events and entities are directly mapped into tables (b)Uses constraints and keys to manage the data</td>
<td>(a) It is only a logical representation (b) Events and entities can introduce redundancy</td>
<td>Data definition control totally depends on the TF-ORM transition rules. Few transition rules cannot be presented automatically, hence needs user interference</td>
</tr>
<tr>
<td>TF-ORM</td>
<td>Bitemporal</td>
<td>×</td>
<td>×</td>
<td>Object-Oriented SQL</td>
<td>Delphi-Query Interface</td>
<td>tf orm</td>
<td>(a) Uses schema versioning on table (b) Classes and their object's roles are timestamped</td>
<td>(a) Schema can easily be mapped to the database (b) Transparency of database is present</td>
<td>Data definition control totally depends on the TF-ORM transition rules. Few transition rules cannot be presented automatically, hence needs user interference</td>
</tr>
<tr>
<td>TSQL2</td>
<td>Valid Time</td>
<td>×</td>
<td>×</td>
<td>Temporal Relational Model</td>
<td>×</td>
<td>SQL</td>
<td>(a) Time is divided into fragments (b) Those fragments record the event occurrences</td>
<td>Allows multiple occurrence of events in same fragments</td>
<td>Tuple can store only future aspects of time and not the past values.</td>
</tr>
<tr>
<td>Relational Database Transformation to Transaction Time Temporal Database</td>
<td>Transaction Time</td>
<td>×</td>
<td>×</td>
<td>PL/SQL and Database Rules</td>
<td>Relational Model</td>
<td>SQL</td>
<td>(a) Table versioning is used (b) Uses an automatic rule based approach for schema transformation</td>
<td>Modification of current data manipulation language is done using proficient rule based approach</td>
<td>Unnecessary trigger execution due to various update and delete operations can introduce delays</td>
</tr>
<tr>
<td>TEER</td>
<td>Bitemporal</td>
<td>×</td>
<td>×</td>
<td>Enhanced Entity-Relationship Model</td>
<td>×</td>
<td>Temporal SQL Extension</td>
<td>(a) Have different relations to focus on regular entity type and weak entity type (b) Uses relations to express multivalued attributes</td>
<td>(a) Easy to understand and expressive (b) Mapping of it in extension of relational model is easy</td>
<td>It does not provide any kind of mapping to implementation model from the TEER diagrams</td>
</tr>
</tbody>
</table>

IV. CONCLUSION AND FUTURE WORK

The research in the area of temporal database has been going on from 20 years or more. In near 70’s Codd proposed the relational model with its relational structure, due to which the main focus was only on the relational model. With time when it was required to add time with the data in an organized way, it results in temporal relational databases. After some decades, need arises to capture real world entities with this time attribute which results in temporal object-oriented data models. A discussion is done on the following points:
• A brief discussion is done on the basic conventional databases including relational database, which is widely used almost everywhere.
• A need of temporal database is reviewed as what problems were prevailing in result of which we require temporal database.
• Various time dimensions or time aspects of the temporal models are discussed and well understood.
• A lot of research has been done on both types of temporal data models: relational and object oriented models, discussing about their complex nature and structural design. But, none of the models satisfies all the desired objectives as an individual model. So, it is recommended to use a suite of data models for the best outcomes.
• There are various kinds of query languages exist that can be used to access or manipulate these data models. All the languages are easy to understand and implement.
• Many temporal data models are discussed from which some are just conceptual models but are equally important as they form a basis for many complex models. But there are some models also, which has there implementation done.

Areas that need to be addressed in near future:
• Attaining a suitable level of performance is a big task that needs to be considered in temporal databases as it records the voluminous amount of data, and sometimes heterogeneous in nature, which will take plenty of time to get processed. There is a need to do analysis of the temporal indexing and temporal join algorithms in order to increase performance.
• As tuple-timestamping increases redundancy in the table, it is needed to manage the increasing size of the table because it will become impractical to store all the data in main memory of one machine. Other storage options can be explored like a distributed environment can be implemented to resolve this.

V. REFERENCES