Journal of Technology Management and Business (ISSN: 2289-7224)
Vol 5, No 2, 2018

ANALYTICS DRIVEN DATA MODEL IN DIGITAL SERVICES

Ng Wai Keat[1]

[1] Axiata Analytics Centre, Axiata Group, Malaysia

*Corresponding E-mail : waikeat.ng@axiata.com

Abstract

Data models are generally applied to construct consolidated abstraction of various rich and different domains of data. In this paper, we focus on the digital services domain in particularly customer related data model and its structure in helping to shape the analytics capabilities. The traditional Entity Relationship Diagram (ERD) is used as the cornerstone of the strategy and further elaboration is made through abstraction to encompass areas in the digital services. A data model is developed to cover both static aspects (customers' profile) and dynamic aspects (customers' behaviour). The foundation of the customer aspect is constructed in classes that represent different types of customer touch points represented as digital footprint which analogize physical activities. The customer dynamic aspects of digital service are modeled with a group of classes where priority is embodied in different associations involving creation and termination of the identified interaction. The suggested data model can be deployed in development of frameworks for customer related applications and enhancement of analytics capabilities.

Keywords: *Data model, Analytics, Digital services*
DOI: https://doi.org/10.30880/jtmb.2018.05.02.001

Received: January 16, 2018   Accepted: January 25, 2018   Published: June 15, 2018

1.0 Introduction

From the digital perspective, data management plays a crucial role to enable useful insights from a unified view of information. This information originates from different data sources that explain from demographic, usages, mobility to subscriptions and consumers' behaviour. In many scenarios, insufficiency of data semantic created vacuum between insights generation and data capability.

A data model describes the data items of a certain part of the perceived reality (business domain) relevant for a specific application or a specific user in a structured way. This model includes the relationships between the data (Maier, 1999). They offer plausible terms for telling our knowledge about the domain. The main benefit of having a data model for a specific domain is for association and distribution of information about the domain and connecting with other domains.

In the expanding world of digital services, various research and studies have been carried out to find out the connecting data sources. However, there is still lack of generally accepted context that encompasses the description and confederacy of digital services data. The motivation for development of a data model on customer centric digital services and analytics started with complexity of managing commonality of data semantics. As data sources grows, finding the unity across multiple data sources becomes challenging and disorganized when analytics is applied into the data to gain insights.

Interoperability is the key driver that spurs the development of a common data model that support the concept of customer analytics driven. In digital services, data come from various sources with multi type of heterogeneity characteristics. The types of heterogeneity are (Sheth, 1999):

- Syntactic: Differences in representation format
- Schematic: Native model that stores data differ in data sources creating structural variance.
- Semantic: Ambiguousness meaning of data
- System: Occurred from different operating system and hardware.

This paper presents the entity-relationship analytics data model for digital services based on a common setting of multiple digital services across several countries. Our data model design adopts the natural view of the real world by combining best practices gathered from digital services in telecommunication space. Despite of various specific localizations in the data, we have created a common data model to represent each entity

2.0 Literature Review

Data model plays a significant role in depicting the higher level of relationship in terms of concepts which assist in defining the underlying lower level concepts. Data model helps in constructing the concepts systematically without lengthy elaboration in complex concepts through formation of relationship between entities. In most scenario, domain data model term is usually used to assert the specific schemes about a domain or a condition in a domain. Upon having the representation of the schemes, the information pertaining the characteristic of the schemes also can be embodied as well.

The key inspiration for developing an analytics data model of digital services is diverse. Firstly, as mentioned in the introduction, there is a strong request to enable uniform semantic representation among different operating digital services in the telecommunication field. There is a huge challenge in knowledge discovery especially with data of massive size and time sensitive in a non-formalized schema set-up. Hence, with data model describes dataset, it will spur the development of robust query and integration capabilities through a few different methods (Goguen) (Bowers, Ludascher, 2004). For understanding customers' behavior in digitized world, the way forward to mine greater insights relies on the harmonization of the attributes in entities.

Secondly, the availability of plenty of data across various data sources creates a pressing need to readjust the wholesomeness of customers' behaviour across time, space and context. In portraying customers' behaviour, there are several circumstances involved. For example, in the scale of engagement for service or product, customer will have states such as learn, engage, act and tell. The footprints of activity are recorded in various data sources and eventually they are converted into an aggregated view to formulate a perspective of customer's interest for a subject. Hence, data model is vital in maintaining the understanding of relationship and involvement of events in relation to the customers' interest.

Lastly, data model creates opportunity of knowledge exchange. Upon reaching the visualization stage, the data model represents an area of knowledge for a domain. The key reason is that the central concept structure of the domain is recorded. Association of terms to concepts and relations in the data model is required to construct a knowledge representation which combines attributes and relationship of objects. Hence, with the knowledge representation, information can be shared to people with similar requirements in specific domain without the need of reproducing the process of analysis.

3.0 Methodology

In information science, data modelling is considered as the process of constructing and organizing data. Both data analysis and data modelling are common in terms of the activity in the ideas and methods of synthesis (Simson,Witt, 2005). We have adopted the 5 steps approach with data coming from various sources that run over repeated cycle as shown in Fig. 1:

Figure 1: The 5 Steps Approach

- Step 1: Selection of schema that have useful outcome that act as common core services across different data sources (various countries). Numerous services are rendered in the digital space and each of the services has multiple touch points. There are touch points which are more crucial for analysis than others. Hence, concised schema selection plays an important role to ensure significant clues about the subscribers' behaviour are captured.
- Step 2: Data collection and analysis on current digital services to detect the shared structure. It is a common practice to create analysis coming from database fields (Wieringa, Jonge, 1995). Similar data can come in different formats and representations. This is especially more visible when data sets are gathered from many digital services across different countries.
- Step 3: Shared data dictionary is established by sanitizing and formatting the relevant fields into meaningful expression. With the knowledge gathered from different data sources, a combined understanding about the fields and definitions is formed. The shared data dictionary is developed based on the combined knowledge of the data sets driven by conformity for semantic interoperability.
- Step 4: All the entities are modelled based on the relationship with each other. Upon classifying the entities, they are map into relationship of classes. The classes are representing the interactions and constraints between entities.
- Step 5: Distribution of the overall schema with the guidelines and training on how to interpret the mapping. Upon completion of the common schema, the description of the schema is documented for dissemination during engagement.

3.1 Data Model

During the initial stage, several studies and reviews are conducted on the existing literature. In our research of data model engineering, we found most of the processes are manual and inventive. There are many light-weighted ontologies deployed that are usually have loose definition of relationship between entities and moderately easier to develop into data model (Mizoguchi). Despite that the domain of digital is relatively budding, elaborative data model is essential to ensure the seamless compatibility between various services offered via the telecommunication channel.

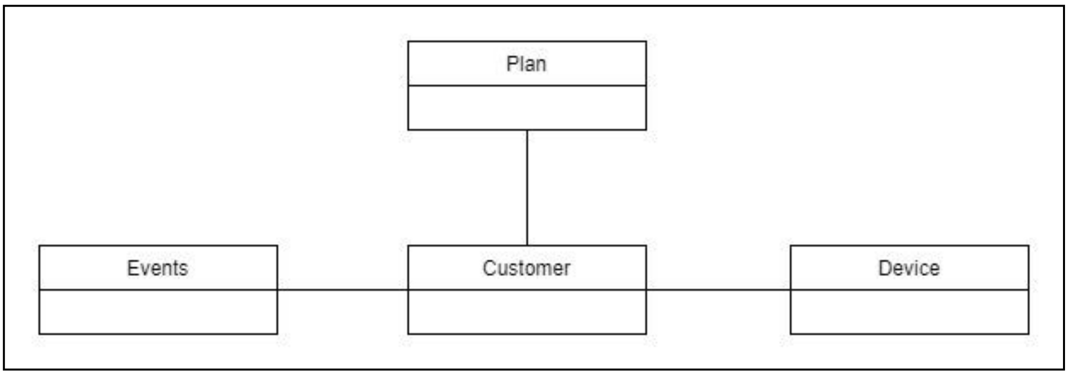

Figure 2: Conceptual analytics data model diagram for Digital Services

The conceptual analytics data model in Fig. 2 depicts the fundamental data source relationship between entity customer with all the associated entities surrounding the activities or states related to the customer. This initial model paves the way to further refine the attributes and articulate the relationship to more layers. Based on the diagram, the customer has association with the package plan, events/ activities in the telco network and the device used to connect to the telco network.

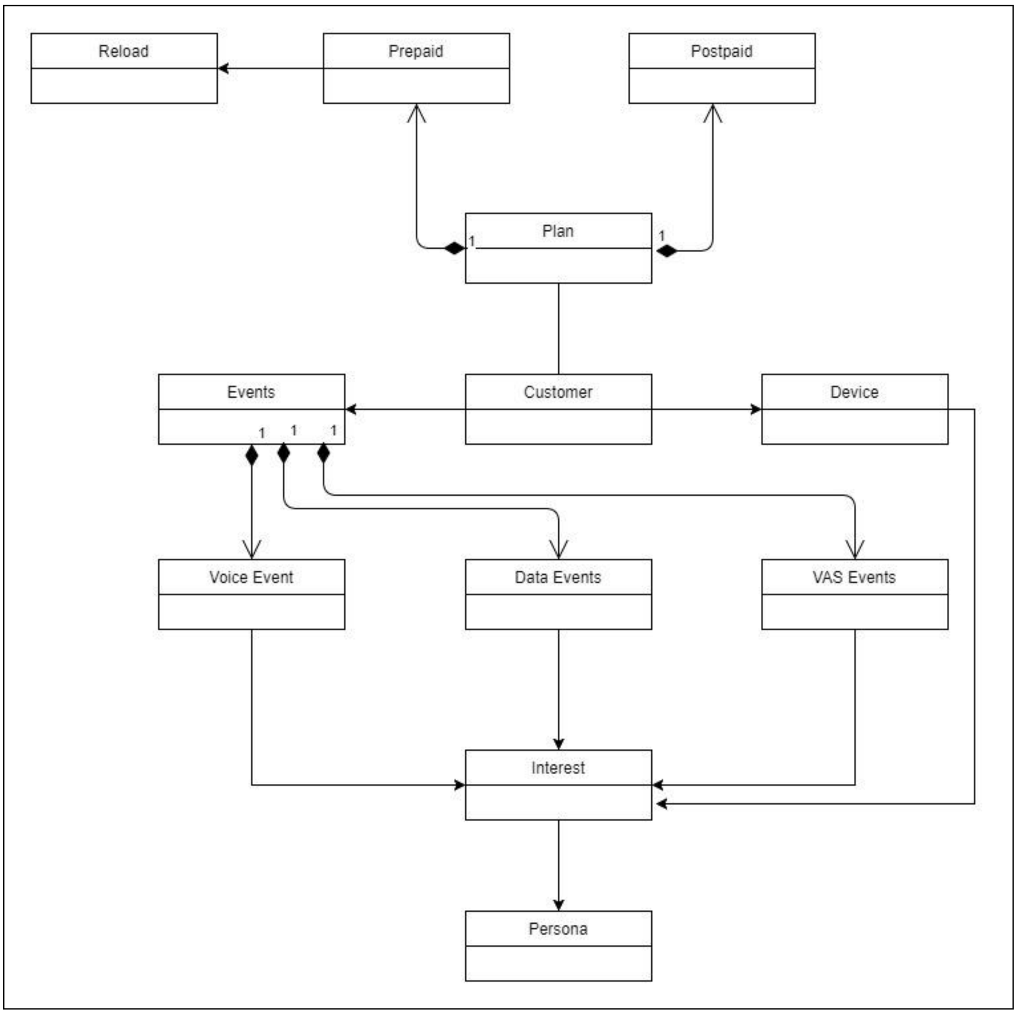

Figure 3: Logical analytics data model diagram for Digital Services

The logical diagram in Fig. 3 shows more details of the relationship of customer to the subscribed services, devices and subscribed plan. The logical model comprises of various telecommunication services that provide footprint of activities both from traditional and digital world. The entities of Prepaid and Postpaid are composite of Plan entity where each has different filtering criteria. In Prepaid entity, it is corresponding with Reload entity using unique identifier of <customer_id>. As for Voice Event, Data Event and VAS Event have data generated as composite of Event entity which corresponds with the Customer entity. Each instance of entry creates <customer_id> corresponding to the record of events. Subsequently, all the events with the device details generate from the same customers are realized into Interest entity where identification of interest is formulated based on rule sets that derived from continuous analysis. Finally, with the grouping of Interest instances, Persona entity is established by capturing similar qualities of activity to establish customers' behaviour. This logical model creates the reflection of the entity relationship structure in the database. Through mapping of the logical model to the detailed entity relationship, the physical implementation can be taken place organically without elaborative description from the logical model.

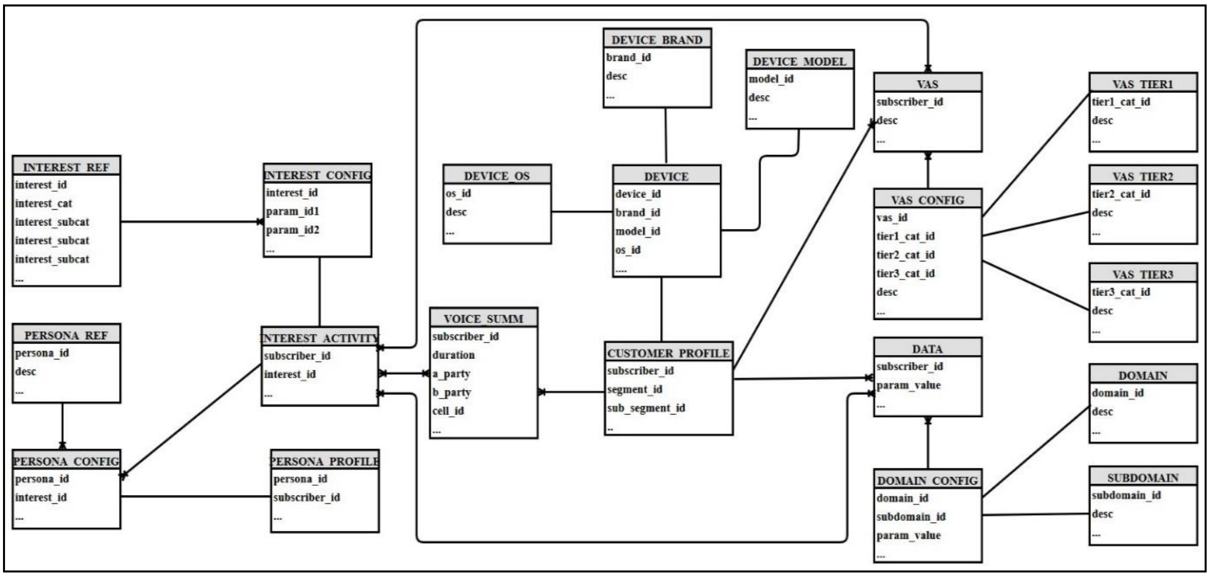

Figure 4: Entity Relationship Diagram for Analytics Driven Data Model for Digital Services

The establishment of entity relationship diagram in the implementation mode is depicted in Fig. 4. Most of the entities are realized through several tables. In most scenario, mapping of entities to database table is not straightforward. For example, a Customer entity could have a property of DeviceType. In the corresponding database table, the Customer_Profile table has a key that reference to Device table that describe the type of device. Deep dive further, one property such as device model may spread into several columns across tables such as device brand and device model number. Usually database tables are normalized to optimize the performance and hence, encapsulate the conceptual meaning.

4.0 Results and Discussions

The continuous in-flow of data that are changing swiftly brought to our attention to baseline the important understandings of entities to a greater extend particularly to ensure common representations. The strategy working forward stands on increasing the standardization and interoperability at syntactic, schematic and semantic level which focuses on digital libraries (Paepcke et al. 1998). Our proposal for an analytics data model for digital services includes the explanation of basic entities. Some of the basic entities are used to create more aggregated representations. In addition, the information involved are heterogeneous and from a wider variation of digital data that can generate new data and information.

At the developing stage, most analyses emphasize on defining the basic and derived entities at the data store level. Fundamentally the approach is taken to ensure the data modelling including the entities and relationships are defined concisely at the earliest phase. The next opportunity available is to extend the data model to encompass the data from curation task, analytics algorithm and components. This would allow more definition of complex entities with inclusion of constraints and scenarios that support numerous relevant services and resources

Future work from this paper would be on derived entities that are explored through consideration of time limitations and sequence of events. During the analysis stage, several data sources are identified as temporal in nature and heavily time sensitive in consuming them for analytics purposes. These data potentially able to manifest into different attributes based on various time frame. Hence, handling of structure and semantic in such trait requires a formal representation of temporal knowledge via top-level data model that describes the most universal categories of temporal entities (Baumann, Loebe, Herre, 2012).

5.0 Conclusions

In this paper, we have presented a proposal for an analytics data model for digital services. The method taken mainly focuses on the customer particularly in understanding relationship between behaviour which led to certain actions. The proposed data model includes definition of basic and derived entities which represent generative information. All the entity relationships are represented using a predefined set of relations. There are still works in progress, where more use cases are reviewed and further refining of the data model is being continuously scrutinized.

Acknowledgement

Special thanks to the Unified Data Model working group in Axiata Digital which started the brainstorming of various ideas to improve the existing representation of data model.

References

Ronald Maier. (1999). Evaluation of Data Modeling. *ADBIS '99 Proceedings of the Third East European Conference on Advances in Databases and Information Systems*, Springer-Verlag, London,.UK.

P. Sheth. (1999). Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics. *Interoperating Geographic Information System*, pp. 5–30.

J. Goguen. Information Integration in Institutions. *In Lawrence Moss (Memorial volume for Jon Barwise, Indiana)*.

S. Bowers & B. Ludascher. (2004). An Ontology-driven Framework for Data Transformation in Scientific Workflows. *Data Integration in the Life Sciences*, Springer, pp. 1-16.

G. Simson G. & G.C. Witt. (2005*)*. Data Modelling Essentials. *Third Edition (Morgan Kaufmann Publications, Elsevier*.

R. J. Wieringa & W. D. Jonge. (1995). *Theory and Practice of Object Systems 1*(2), pp. 101-114.

R. Mizoguchi. *Tutorial on ontological engineering - part 3: Advanced course of ontological engineering* (New Generation).

Paepcke, C-C. K. Chang, T. Winograd & H. Garcia-Molina. (1998). *Communications of the ACM 41*(4), pp. 33-42

Baumann R, Loebe F, Herre H. (2012). Ontology of time in GFO. *Formal Ontology in Information Systems, Proceedings of the Seventh International Conference,* The NetherlandsIOS Press, Amsterdam, pp. 293-306.