Chapter 6
A Disciplined Approach to Temporal Evolution and Versioning Support in JSON Data Stores

Safa Brahmia
University of Sfax, Tunisia

Zouhaier Brahmia
https://orcid.org/0000-0003-0577-1763
University of Sfax, Tunisia

Fabio Grandi
https://orcid.org/0000-0002-5780-8794
University of Bologna, Italy

Rafik Bouaziz
University of Sfax, Tunisia

ABSTRACT

The JSON Schema language lacks explicit support for defining time-varying schemas of JSON documents. Moreover, existing JSON NoSQL databases (e.g., MongoDB, CouchDB) do not provide any support for managing temporal data. Hence, administrators of JSON NoSQL databases have to use ad hoc techniques in order to specify JSON schema for time-varying instances. In this chapter, the authors propose a disciplined approach, named Temporal JSON Schema (τJSchema), for the temporal management of JSON documents. τJSchema allows creating a temporal JSON schema from (1) a conventional JSON schema, (2) a set of temporal logical characteristics, for specifying which components of a JSON document can vary over time, and (3) a set of temporal physical characteristics, for specifying how the time-varying aspects are represented in the document. By using such characteristics to describe temporal aspects of JSON data, τJSchema guarantees logical and physical data independence and provides a low-impact solution since it requires neither updates to existing JSON documents nor extensions to related JSON technologies.

DOI: 10.4018/978-1-5225-8446-9.ch006

INTRODUCTION

JSON (IETF, 2017) is a standard format for interchanging data between all programming languages (EMCA, 2017). On the Web, it is usually used for structuring and sending data from a server to a client or vice versa, so that they could be displayed on a Web page or processed by a Web service. In the database context, JSON is also a new database model for NoSQL data (Cattell, 2010; Tiwari, 2011; Pokorný, 2013; Davoudian *et al.*, 2018), whatever they are structured or semi-structured. Due to the dynamic nature of modern computer science applications (e.g., social networks, IoT, cloud computing, smart cities), JSON documents that are exploited by these applications —like other application components such as scripts' source code and graphical user interfaces— evolve over time to reflect changes that occur in user requirements and in the modeled reality. Moreover, several applications (like mobile, GIS, e-health and e-government applications) necessitate keeping track of JSON data evolution with regard to time, requiring time-varying JSON documents to be represented, stored and retrieved.

However, the current JSON format and state-of-the-art JSON NoSQL database systems (e.g., MongoDB, CouchDB, DocumentDB, Couchbase Server, MarkLogic, OrientDB, RethinkDB, Riak, Elasticsearch) and JSON management tools do not provide any built-in support for temporal JSON documents, despite the steady interest for temporal and evolution aspects among researchers and practitioners (Cuzzocrea, 2015). In particular, also the latest JSON Schema specification (IETF, 2018) lacks explicit support for time-varying data, both at schema and instance levels. Thus, JSON NoSQL database administrators (JNoDBA) (i.e., any person who is in charge of the maintenance of either a JSON NoSQL database or a JSON repository) must use ad hoc techniques when there is a need, for example, to specify a JSON Schema for time-varying JSON documents.

According to what precedes, we think that if we would like to efficiently handle JSON document evolution over time and to allow temporal queries to be executed on time-varying JSON documents, a comprehensive temporal JSON NoSQL database management system is required. To this purpose, we present in this chapter an approach, called τJSchema, for managing temporal JSON documents through the use of a temporal JSON schema. In fact, we want to introduce with τJSchema a disciplined approach to the temporal extension of JSON Schema, similar to what has been done with the τXSchema approach (Currim *et al.*, 2004; Snodgrass *et al.*, 2008) to XML Schema (W3C, 2004) management. τXSchema is a well-known approach in the temporal XML database (Brahmia & Bouaziz, 2008; Dyreson & Grandi, 2018) community, which consists of a data model equipped with a suite of tools for managing temporal XML documents.

Since it is proposed as a τXSchema-like approach, τJSchema allows constructing a temporal JSON schema from a conventional (i.e., non-temporal) JSON schema and a set of temporal logical and temporal physical characteristics. Temporal logical characteristics identify which components of a JSON document can vary over time; temporal physical characteristics specify how the time-varying aspects are represented in the document. By using both temporal schema and temporal characteristics to introduce temporal aspects in the conventional JSON NoSQL world, our framework (i) guarantees logical and physical data independence (Burns *et al.*, 1986) for temporal JSON schemas and (ii) provides a low-impact solution since it requires neither modifications of existing JSON documents, nor extensions to all related JSON technologies (including the JSON format, the JSON Schema specification, existing JSON/JSON Schema tools and APIs, and JSON NoSQL databases).