



Contribution ID: 210

Type: Oral

Using Graph Databases in HEP

Thursday 7 November 2019 15:15 (15 minutes)

Data in HEP are usually stored in tuples (tables), trees, nested tuples (trees of tuples) or relational (SQL-like) databases, with or without a defined schema. But many of our data have a graph structure without a schema, or with a weakly imposed schema. They consist of entities with relations, some of which are known in advance, but many are created later, as needs evolve. Such structures are not well covered by relational (SQL) databases. We don't only need the possibility to add new data with predefined relations. We also need to add new relations. Graph databases have existed for a long time but they have only recently matured, mostly thanks to Big Data and Machine Learning. There are now very good implementations and de-facto standards widely available.

The difference between SQL and Graph Database is similar to the difference between Fortran and C++. On one side, a rigid system, which can be very optimized. On the other side, a flexible dynamical system, which allows expressing of complex structures. Graph Database is a new synthesis of object-oriented and relational databases. It allows the expression of a web of objects without the volatility of the object-oriented world. It captures only essential relations, it doesn't keep a complete object dump. Migration to Graph Database means moving structure from data to code, together with migration from imperative to declarative semantics (things don't 'happen', they 'exist'). Data, stored in a structured Graph Database also allows new ways of easy analysis with the help of the Non-Euclidian Machine Learning methods. Those methods are not based on the geometrical structure of the problem domain, but on its topology. This makes the Graph Database structure particularly useful for such an approach.

This presentation will describe the basic principles of the Graph Database together with an overview of existing standards and implementations. The usefulness and usability will be demonstrated for the concrete example of the ATLAS EventIndex in two approaches - as the full storage (all data are in the Graph Database) and meta storage (a layer of schema-less graph-like data implemented on top of either NoSQL or relational storage). The usability, the interfaces with the surrounding framework and the performance of those solution will be discussed. The possible more general usefulness for generic experiments' storage will be also discussed. Some examples of using Graph-like data for simple Machine-Learning processing will be also shown.

Consider for promotion

No

Author: HRIVNAC, Julius (Centre National de la Recherche Scientifique (FR))

Presenter: HRIVNAC, Julius (Centre National de la Recherche Scientifique (FR))

Session Classification: Track 4 –Data Organisation, Management and Access

Track Classification: Track 4 –Data Organisation, Management and Access