Hybrid: A Large-scale Online In-memory Image Analytics Engine

Michael Gubanov
Department of Computer Science
University of Texas, San Antonio
1 UTSA Circle,
San Antonio, TX, 78249
mikhail.gubanov@utsa.edu

ABSTRACT

Analytical image/video processing tasks such as scene/face/activity recognition are historically performed outside most relational database management systems. Relational engines are optimized for relational data, hence naturally have weaker support for non-relational data such as images or video.

Hybrid, a high-velocity in-memory analytics engine, supports advanced access capabilities to both image/video contents and structured data via the same language - SQL. Briefly, the user would be able to query both relational (rows and columns of a table) and video contents (objects, activities, scene attributes) in a single SELECT statement, hence performing an analytical task on both types of data without an expensive ETL (extract-transform-load) process.

In Hybrid, images/video are stored as matrices in a relational table column and support a subset of linear algebra operators in our hybrid SQL to manipulate them. This approach differs from classical, relational storage, where images/video are stored as BLOBS and access to this type through SQL is limited, because it is stored as an unstructured binary object. The user has to develop an external user-defined function (e.g. in C++) and call it later in order to meaningfully use BLOB data.

Hybrid provides access to both video/image and structured data from the same SQL statement, which enables “hybrid analytics” - analytics on heterogeneous, not only relational data. We demonstrate interactive complex analytics (activity and pattern recognition) in SQL using both matrix and tabular data at scale.

8th Biennial Conference on Innovative Data Systems Research (CIDR ’17) January 8-11, 2017, Chaminade, California, USA.