Bridging the gap between data lakes and RDBMSs - Efficient query processing with Parquet

In the age of massive data, time-intensive loading phases make databases less viable for data exploration tasks. Still, the highly optimized query engines of database systems are greatly beneficial for the performance of data analysis tasks.

With our research, we want to bridge this gap and provide paramount analytical performance without the need of static data loading.

Our approach enables the integration of Parquet files — one of the most used columnar file formats in data lakes — into the data processing pipeline of a database system in a convenient way. We allow end-users to benefit from the database system performance without a costly and time-consuming loading phase.

Type of Poster

A solution

Primary author: REY, Alice

Presenter: REY, Alice

Session Classification: Poster

Track Classification: Poster