HyPerInsight: Data Exploration Deep Inside HyPer

Nina Hubig, Linnea Passing, Maximilian E. Schüle, Dimitri Vorona, Alfons Kemper, Thomas Neumann
{hubig, passing, schuele, vorona, kemper, neumann}@in.tum.de

Apriori
k-Means
DBscan
k- Modes
NaiveBayes
PageRank

Data mining operators

Airlines
Airports
Routes

OpenFlights dataset

Airline Prediction
Airplane Prediction
Airport Clustering/Rank

SQL

Operators + Lambdas

+ Leverages sophisticated SQL optimizer
+ Reduces memory footprint compared to recursive CTEs
  – User has to implement complex algorithms
  – Invariants and assumptions are opaque to the optimizer
  – Non-standard SQL extension
+ Seamless integration with SQL
+ Fully optimizable
+ Allows re-using of large parts of the query execution
  – Requires deep knowledge of database implementation
  – Every variant and parametrization has to be implemented
  This is solved by introducing lambda functions