

Coral: A SQL translation and rewrite engine for modern data lakes

(Invited Talk)

Walaa Eldin Moustafa

Senior Staff Software Engineer

LinkedIn

wmoustafa@linkedin.com

Abstract

In this talk, we present Coral, a framework for achieving logic interoperability between SQL engines. Coral defines a standard intermediate representation (IR) that is used to express different SQL dialects, relational languages, and query plans. Coral implements a set of adapters to convert between various input and output representations by mapping their semantics to and from Coral IR. Currently Coral supports a number of conversions between Hive, Spark, Trino (or Presto) dialects of SQL. Further, we discuss Coral applications such as making views portable across execution engines, and its future extensions to support common query optimizations such as materialized views selection and substitution, SQL pushdown, and incremental compute.

Biography

Walaa Eldin Moustafa is a Senior Staff Software Engineer at LinkedIn, where he works on building big data infrastructure and solutions for enabling unified and performant data processing systems across different compute engines, storage representations, and language APIs. Walaa holds a PhD degree in Computer Science from the University of Maryland at College Park. He has co-authored a number of database publications at various database conferences including SIGMOD, ICDE, and IEEE Big Data in topics that focus on modern applications of relational and deductive database management systems, such as graph query processing, machine learning, data integration, and probabilistic databases.