Algebricks: A Data Model-Agnostic Compiler Backend for Big Data Languages

Vinayak Borkar, Yingyi Bu, E. Preston Carman, Jr., Nicola Onose, Till Westmann, Pouria Pirzadeh, Michael J. Carey, Vassilis J. Tsotras
UC Irvine, X15 Software, Inc., UC Riverside, Oracle Labs


The Algebricks Framework
- Set of logical operators
- Set of physical operators
- Rewrite rule framework
- Set of generally useful rewrite rules
- Metadata provider API exposing metadata (catalog) info to Algebricks
- Mapping of physical operators to Hyracks runtime operators and connectors

Hivyesterix Example (HiveQL)
```
select sum(l_extendedprice) as revenue
from lineitem
where l_shipdate >= '1994-01-01'
and l_shipdate < '1995-01-01'
and l_discount > 0.05
and l_discount <= 0.07
and l_quantity < 24;
```

See Our Paper For More Information

- Design, implementation, use cases, and performance evaluation of Algebricks
- Algebricks, Apache AsterixDB, and Apache VXQuery all built using Algebricks
- Algebricks is available in open source under the Hyracks repository of AsterixDB
- Future thoughts: Add cost-based optimization and enhance the interaction between Algebricks and Hyracks to support dynamic query re-optimization

While Algebricks is based on Hyracks, similar ideas could be used by other Big Data stacks (e.g., Spark, Flink, or Terz)

Algebricks is available in open source under the Hyracks repository of AsterixDB (https://github.com/apache/incubator-asterixdb-hyracks)

We hereby invite other Big Data researchers to download and try the system!
While Algebricks is based on Hyracks, similar ideas could be used by other Big Data stacks (e.g., Spark, Flink, or Terz)