

Apache Calcite Overview



Julian Hyde
Kylin Meetup (eBay, San Jose)
December 4th, 2014



Apache Calcite

Apache incubator project since May, 2014

Originally named Optiq

Query planning framework

Relational algebra, rewrite rules, cost model

Extensible

Packaging

Library (JDBC server optional)

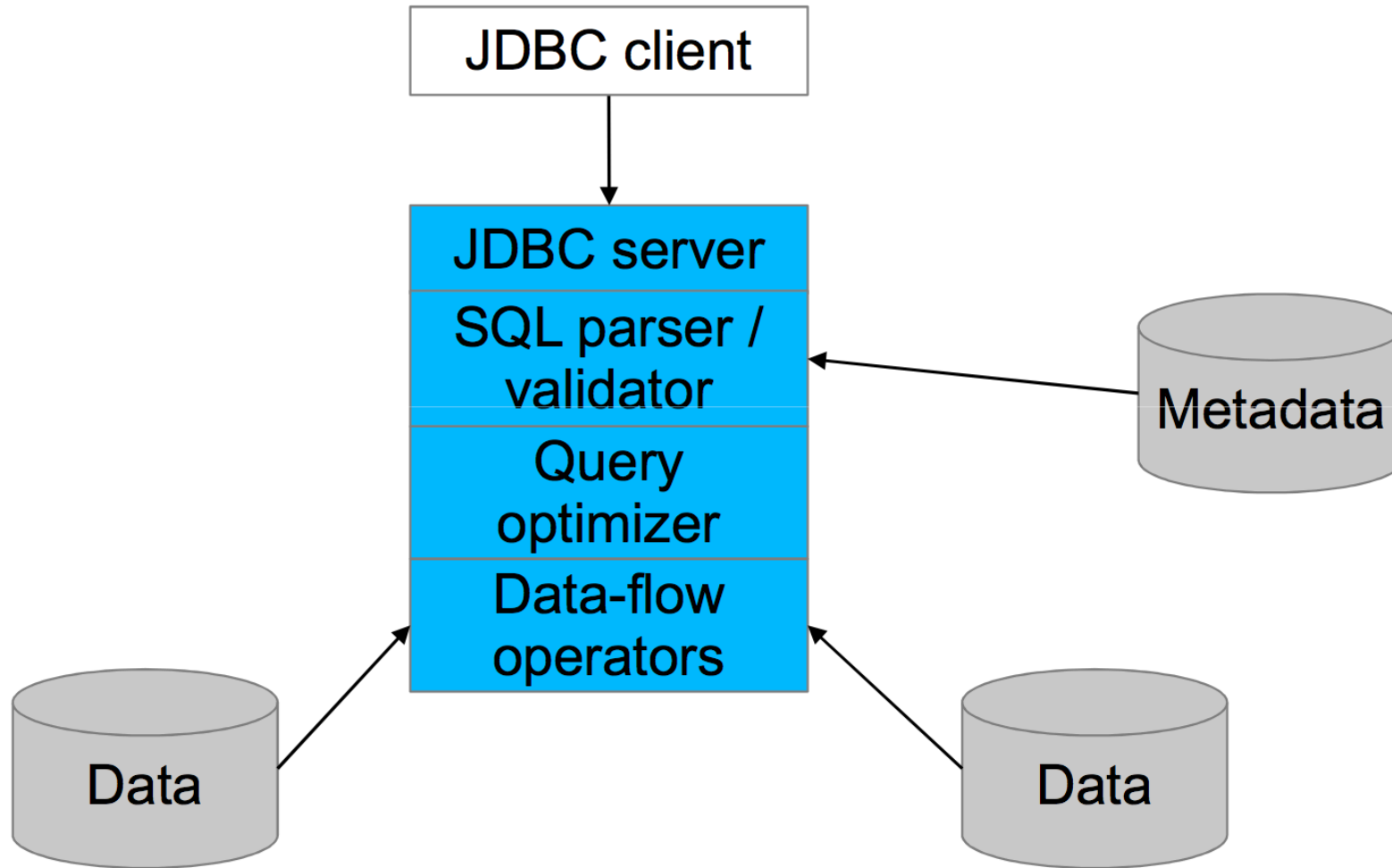
Community-authored rules, adapters

Adoption

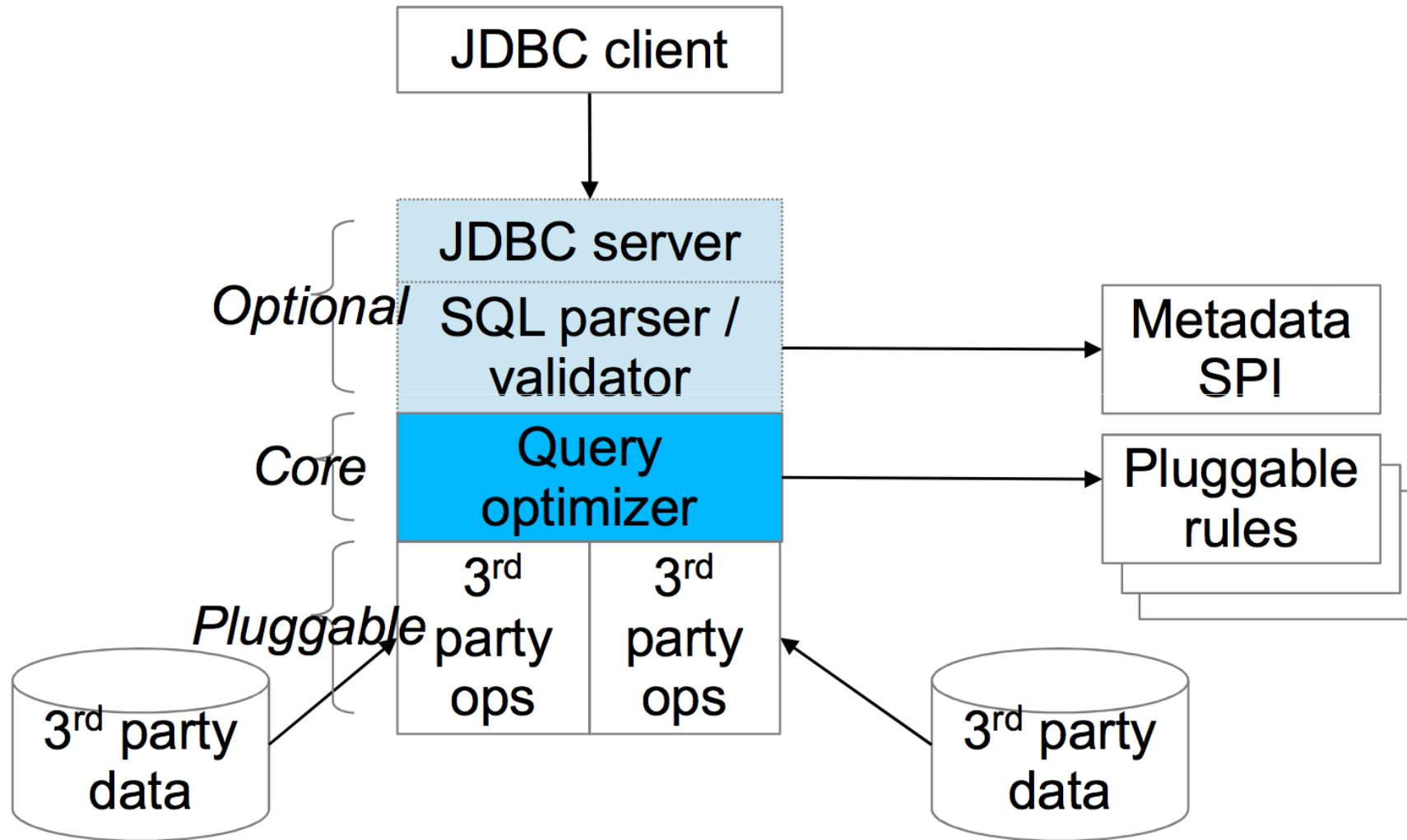
Embedded: Lingual (SQL interface to Cascading), Apache Drill, Apache Hive, Apache Kylin

Adapters: Splunk, Spark, MongoDB, JDBC, CSV, JSON, Web tables, In-memory, Phoenix

Conventional DB architecture

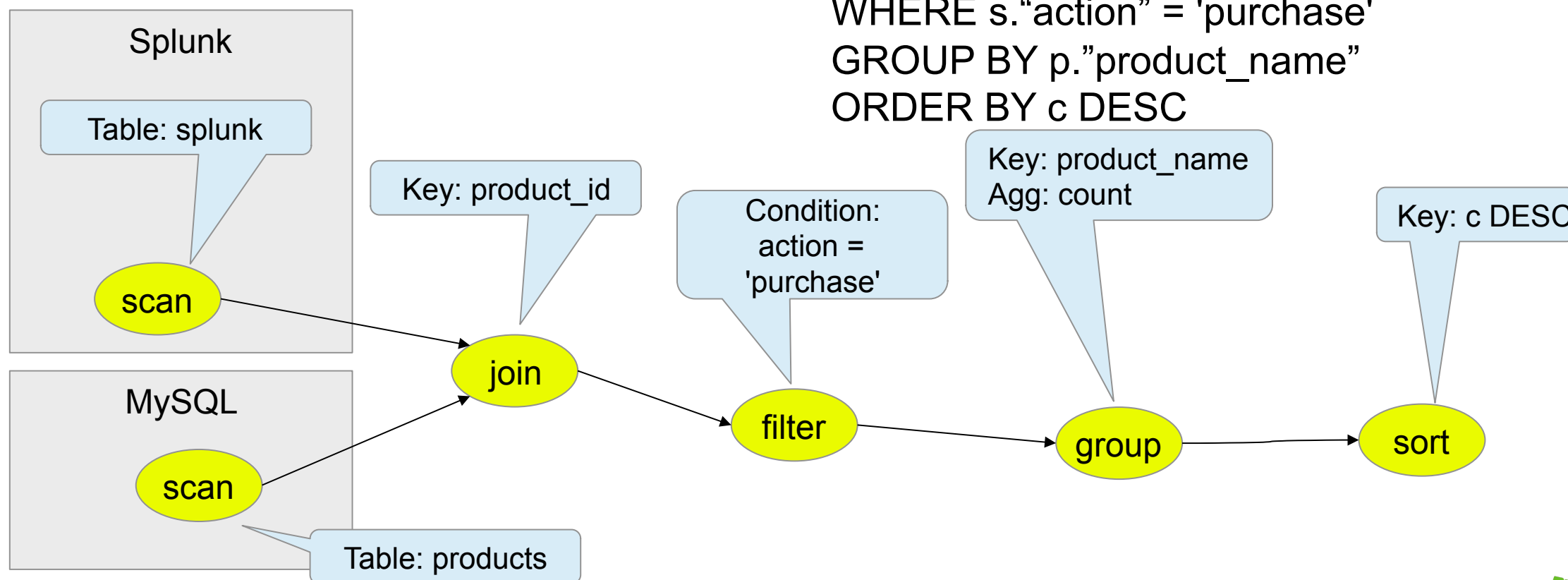


Calcite architecture



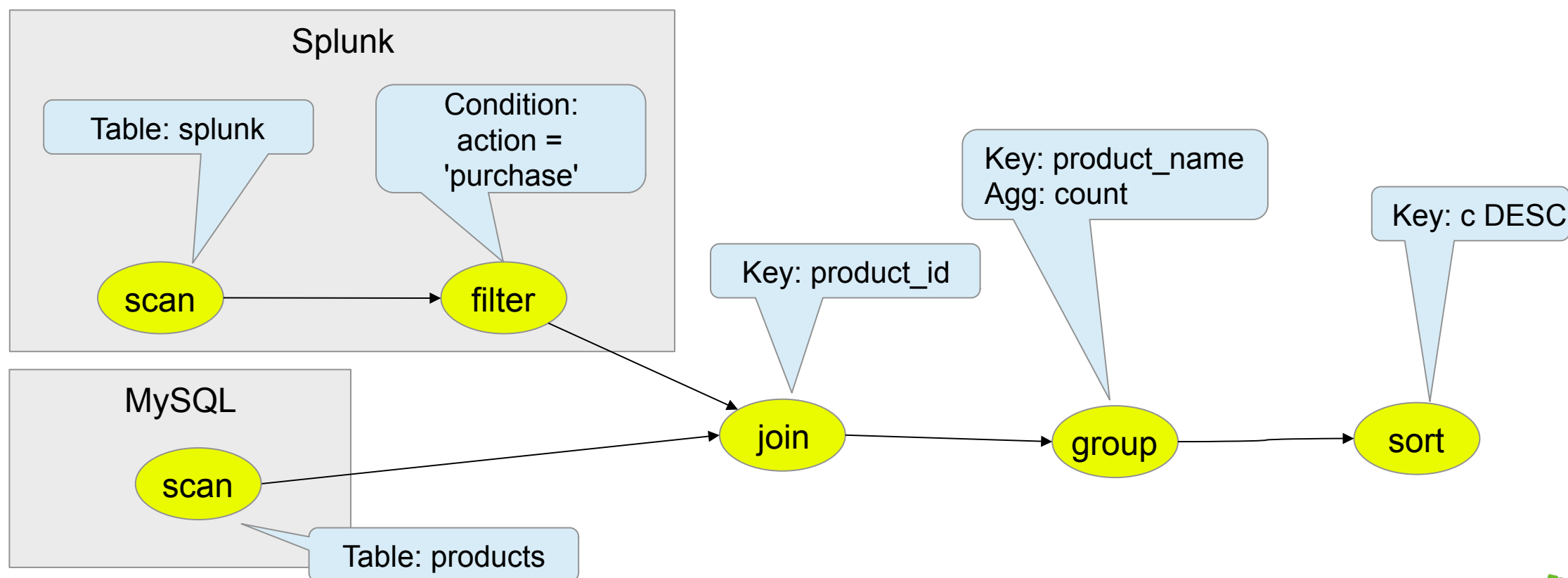
Expression tree

```
SELECT p."product_name", COUNT(*) AS c
FROM "splunk"."splunk" AS s
      JOIN "mysql"."products" AS p
      ON s."product_id" = p."product_id"
WHERE s."action" = 'purchase'
GROUP BY p."product_name"
ORDER BY c DESC
```

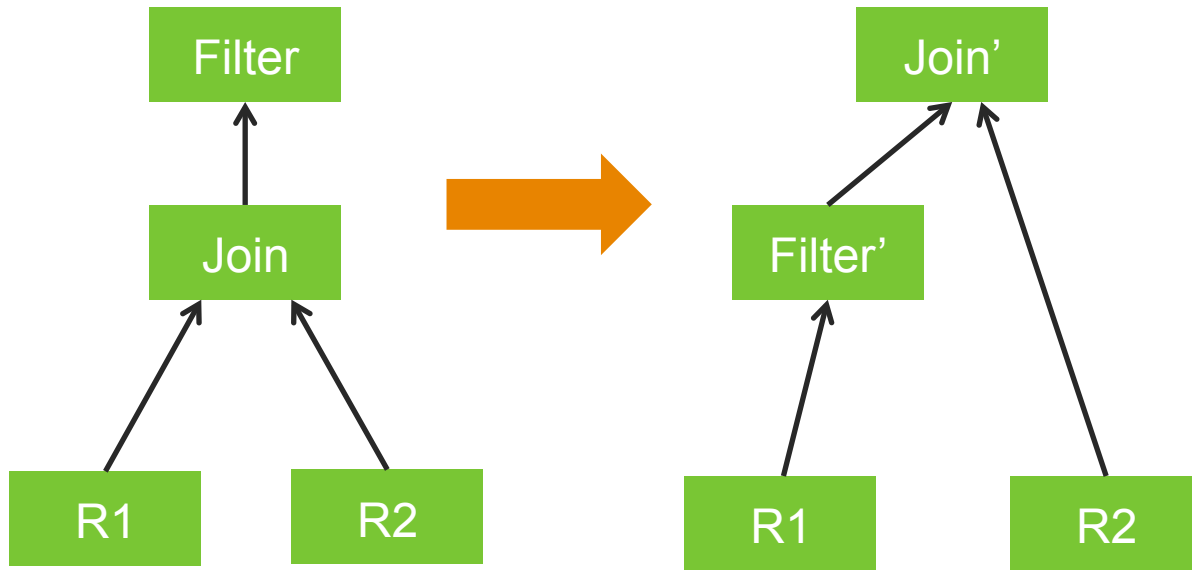


Expression tree (optimized)

```
SELECT p."product_name", COUNT(*) AS c
FROM "splunk"."splunk" AS s
    JOIN "mysql"."products" AS p
    ON s."product_id" = p."product_id"
WHERE s."action" = 'purchase'
GROUP BY p."product_name"
ORDER BY c DESC
```



Defining a rule



```
class FilterIntoJoinRule extends RelOptRule {
    public FilterIntoJoinRule() {
        super(
            operand(Filter.class,
                operand(Join.class, any())));
    }

    public void onMatch(RelOptRuleCall call) {
        Filter filter = call.rel(0);
        Join join = call.rel(1);
        Filter newFilter = ...;
        Join newJoin = ...;
        call.transformTo(newJoin);
    }
}
```

Calcite – APIs and SPIs

Relational algebra

RelNode (operator)

- Scan
- Filter
- Project
- Union
- Aggregate
- ...

RelDataType (type)

RexNode (expression)

RelTrait (physical property)

- RelConvention (calling-convention)
- RelCollation (sortedness)
- TBD (bucketedness/distribution)

Transformation rules

RelOptRule

- MergeFilterRule
- PushAggregateThroughUnionRule
- RemoveCorrelationForScalarProjectRule
- 100+ more

Unification (materialized view)

Column trimming

SQL parser

SqlNode

SqlParser

SqlValidator

Metadata

Schema

Table

Function

- TableFunction
- TableMacro

Cost, statistics

RelOptCost

RelOptCostFactory

RelMetadataProvider

- RelMdColumnUniqueness
- RelMdDistinctRowCount
- RelMdSelectivity

JDBC driver

Thank you!



@julianhyde

<http://calcite.incubator.apache.org/>