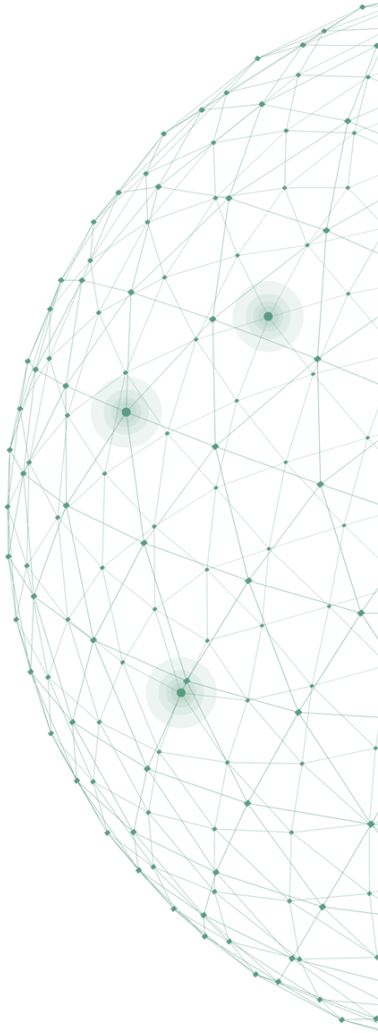
openLooKeng ClickHouse Connector 开发经验

何正杰

openLooKeng

https://openlookeng.io

目录



openLooKeng

https://openlookeng.io

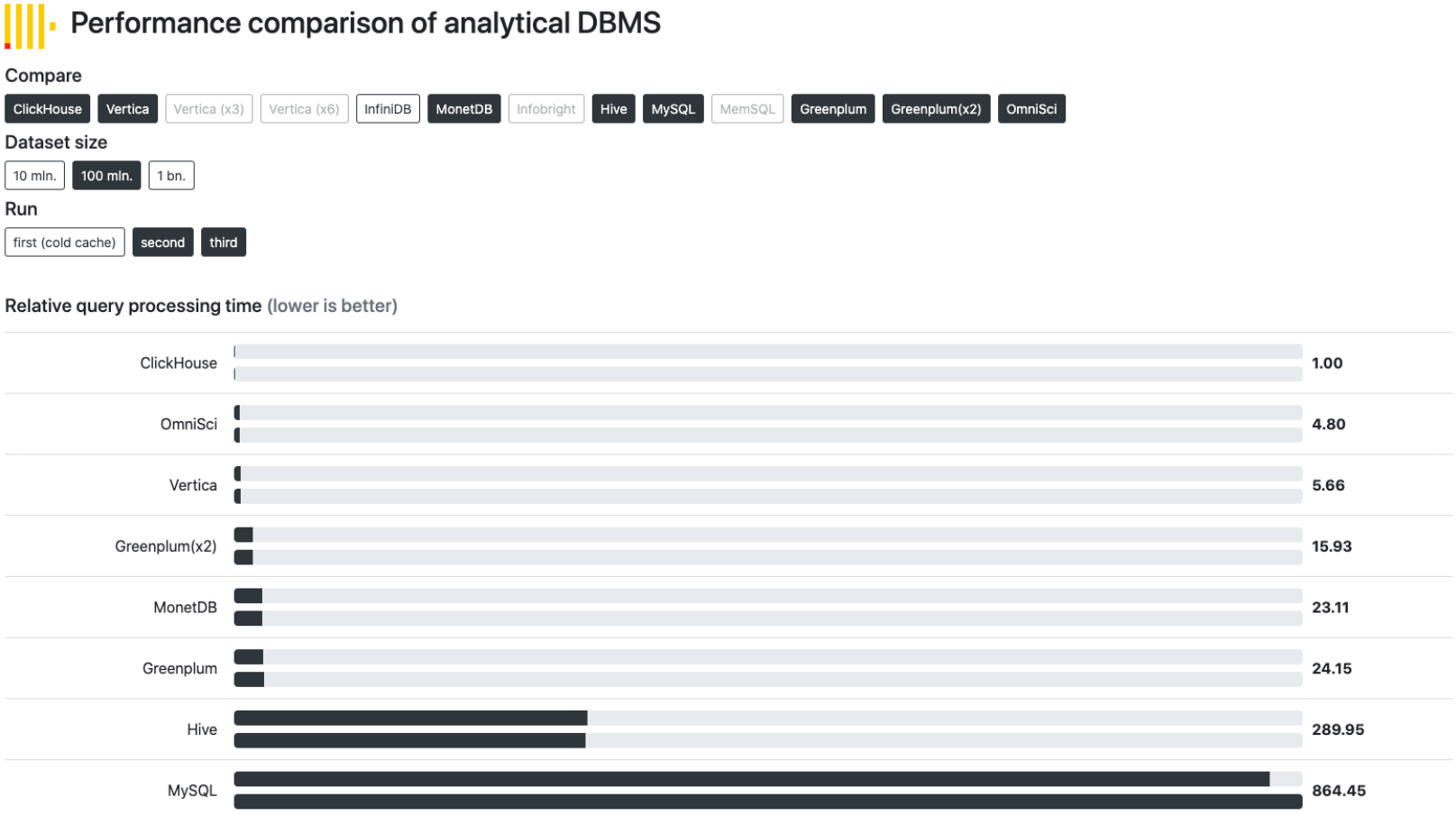
ClickHouse 简介

- ClickHouse是Yandex开源的一个用于联机分析(OLAP)的列式数据库管理系统(DBMS)
- 在多个测试结果中，ClickHouse表现出了比同类可比较产品更优的性能

- 注意点
  - 不支持事务
  - 对于更新和删除支持较低
  - 多表Join的性能不如单表查询
  - 非完全兼容ANSI SQL标准
  - 高并发支持有限，官方建议qps为100

Ref: https://clickhouse.tech/#independent-benchmarks

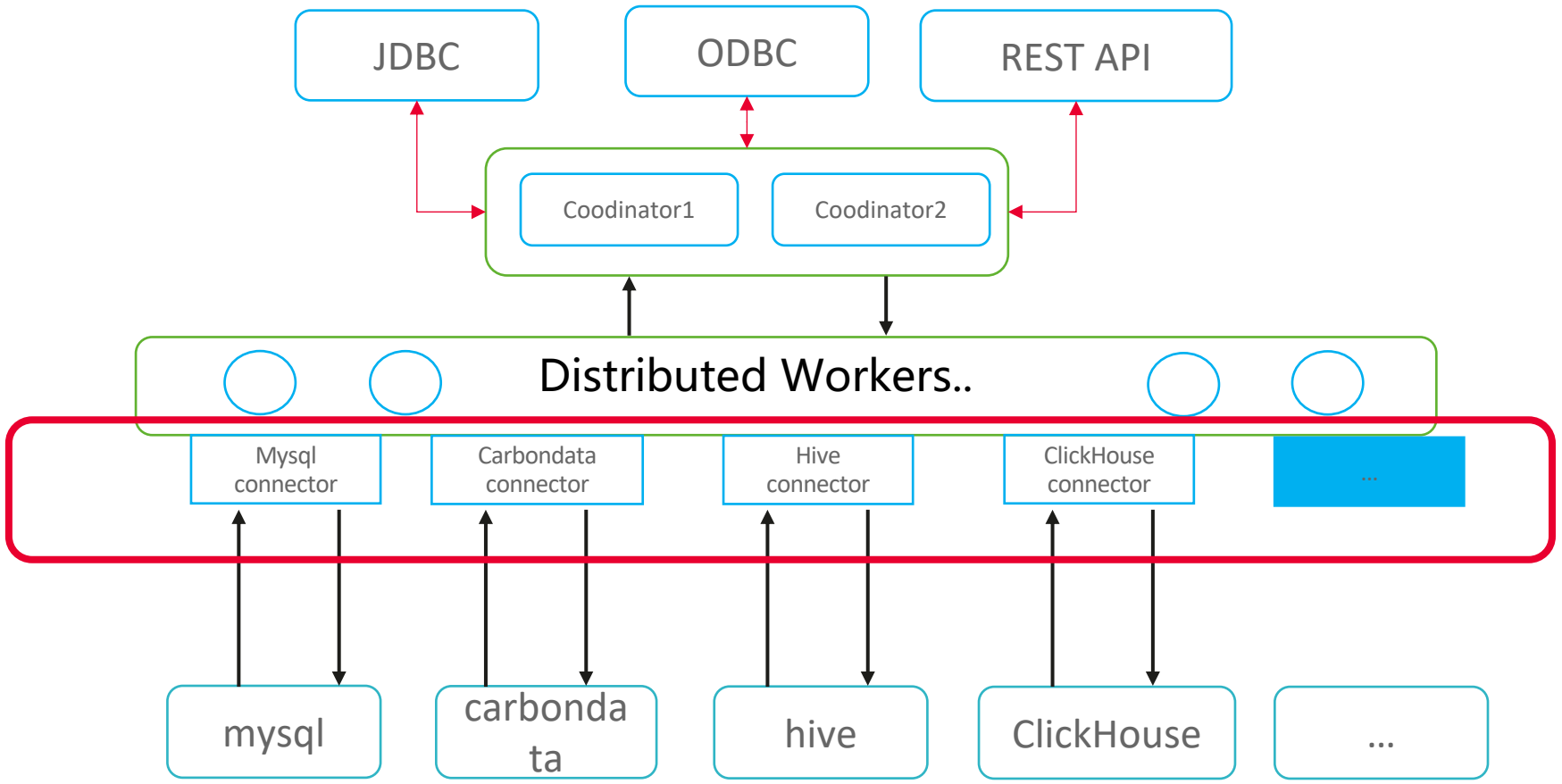
ClickHouse 简介



openLooKeng

https://openlookeng.io

基本架构

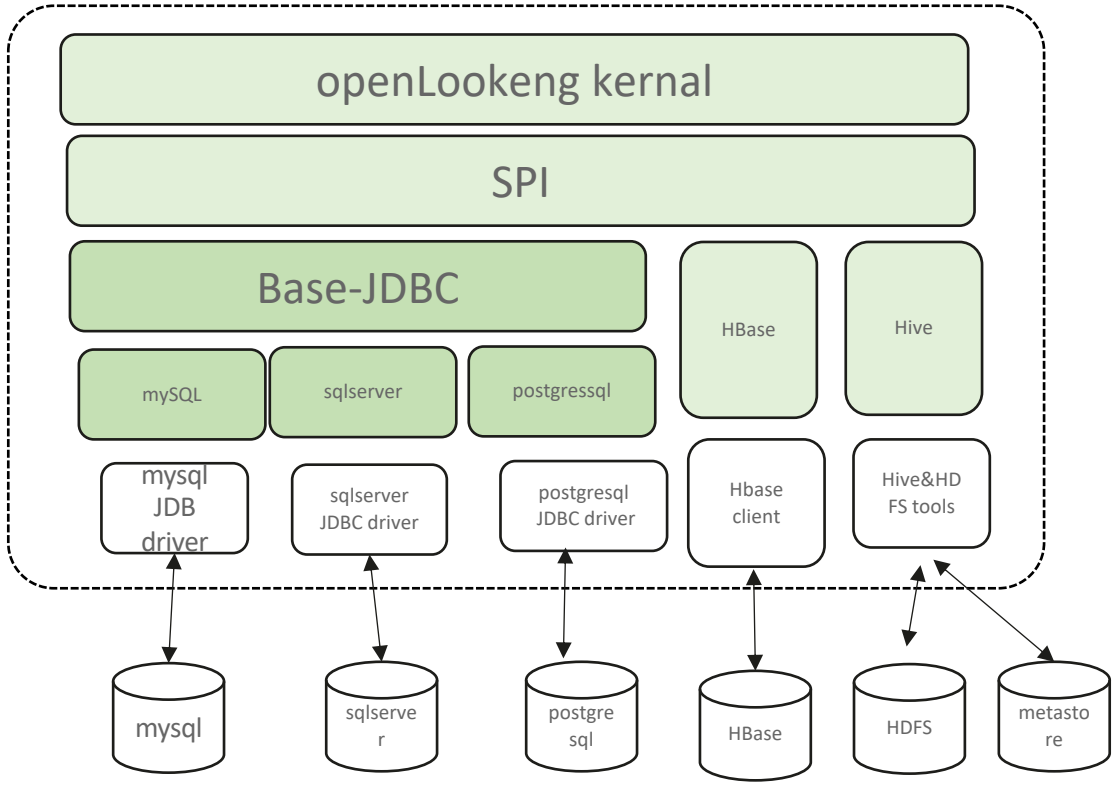


openLooKeng

https://openlookeng.io

ClickHouse connector 适配开发

- 基于Base JDBC开发CH connector

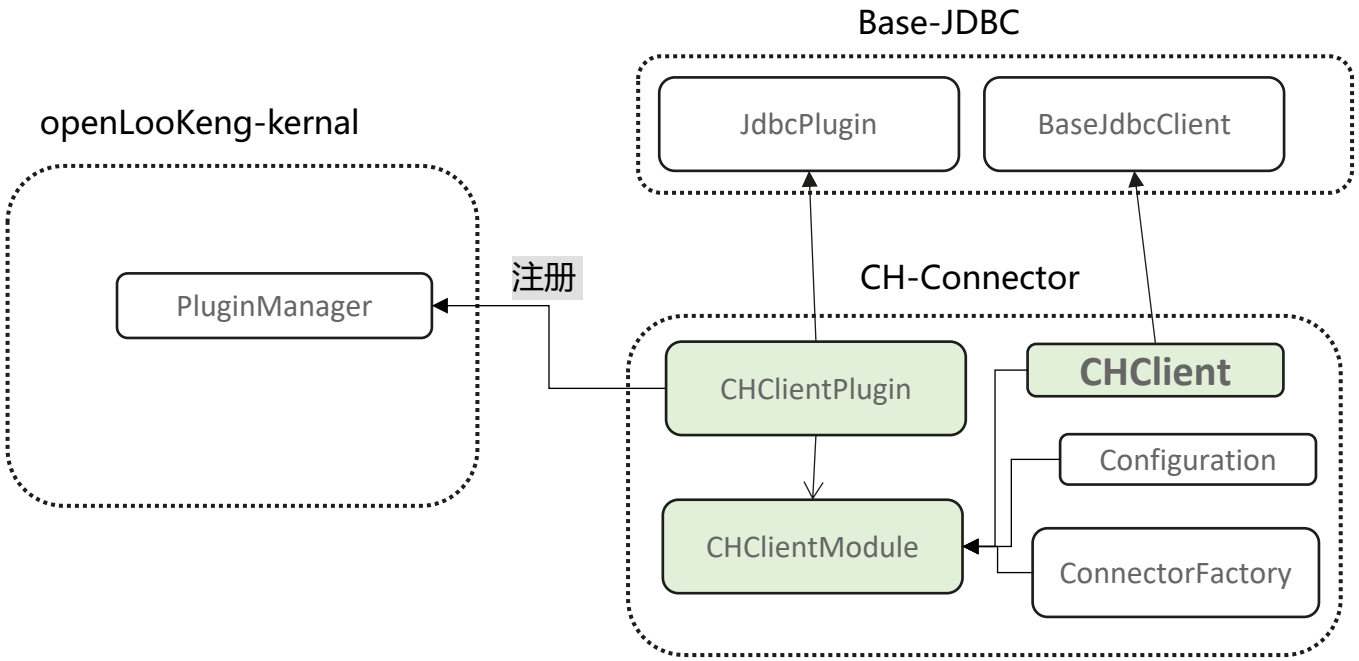


openLookeng

https://openlookeng.io

ClickHouse connector 适配开发

- 基于Base JDBC开发CH connector



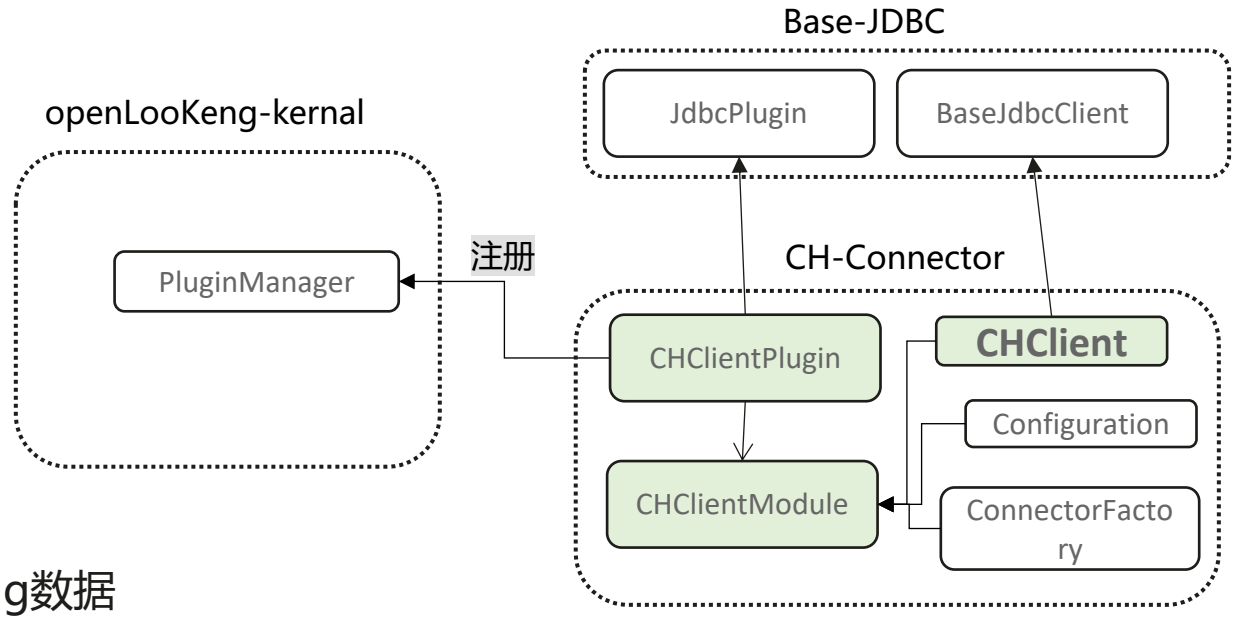
How to add a connector in openLooKeng *https://www.bilibili.com/video/BV1it4y1e7cc*

ClickHouse connector 适配开发

- io.prestosql.plugin.jdbc.JdbcClient
  - > 提供元数据的增删改查等
  - > 提供数据类型转换接口

- 元数据的增删改查
  - > getSchemaNames、listSchemas、getTableNames、getTables、getColumns等

- 数据类型转换接口
  - > 元数据取回时，需要进行数据源类型转化为openLooKeng数据类型-已经有标准实现：toPrestoType
  - > 数据写入，数据需要从openLookeng内核类型转为数据源数据类型openLooKeng数据类型：toWriteMapping



How to add a connector in openLooKeng *https://www.bilibili.com/video/BV1it4y1e7cc*

openLooKeng

https://openlookeng.io

ClickHouse connector 实现

- 数据类型转换：ClickHouse <-> openLooKeng
- 部分类型尚未支持

ClickHouse类型	openLooKeng类型	说明
Int8	TINYINT	
Int16	SMALLINT	
Int32	INTEGER	
Int64	BIGINT	
float32	REAL	
float64	DOUBLE	
DECIMAL(P,S)	DECIMAL(P,S)	
DECIMAL32(S)	DECIMAL(P,S)	
DECIMAL64(S)	DECIMAL(P,S)	
DECIMAL128(S)	DECIMAL(P,S)	
String	VARCHAR	
DateTime	TIMESTAMP	
Fixedstring(N)	CHAR	
UInt8	SMALLINT	
UInt16	INT	
UInt32	BIGINT	
UInt64	BIGINT	
Int128,Int256,UInt256	不涉及	

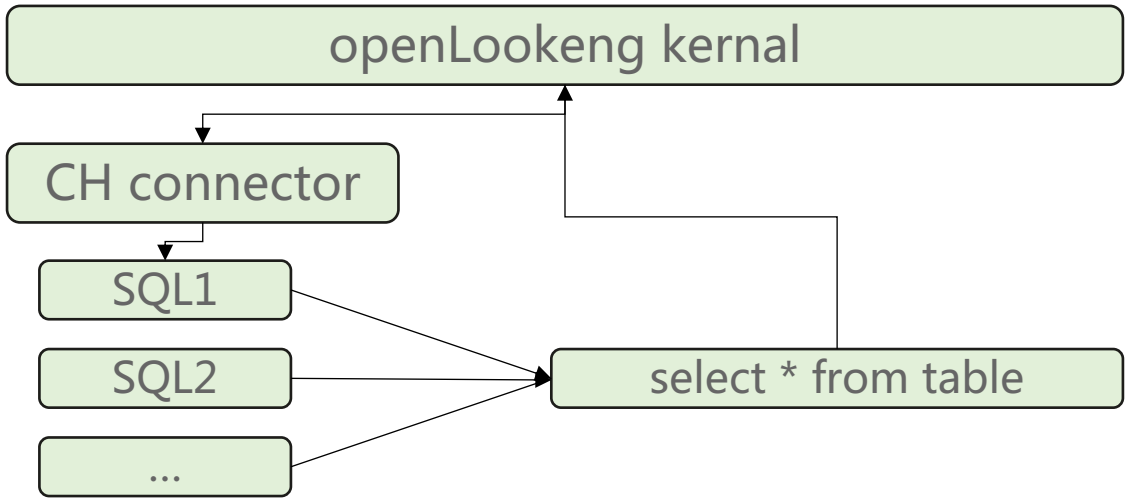
openLooKeng类型	ClickHouse数据库类型	说明
BOOLEAN	Int8	
TINYINT	Int8	
SMALLINT	Int16	
INTEGER	Int32	
BIGINT	Int64	
REAL	float32	
DOUBLE	float64	
DECIMAL(P,S)	DECIMAL(P,S)	
varchar	String	
varchar(n)	String	
CHAR(n)	FixedString(n)	
VARBINARY	String	
JSON	不涉及	
DATE	DATE	
TIME	不涉及	
TIME WITH TIME ZONE	不涉及	
TIMESTAMP	DateTime	
TIMESTAMP WITH TIME ZONE	不涉及	

openLooKeng

https://openlookeng.io

ClickHouse connector 适配开发

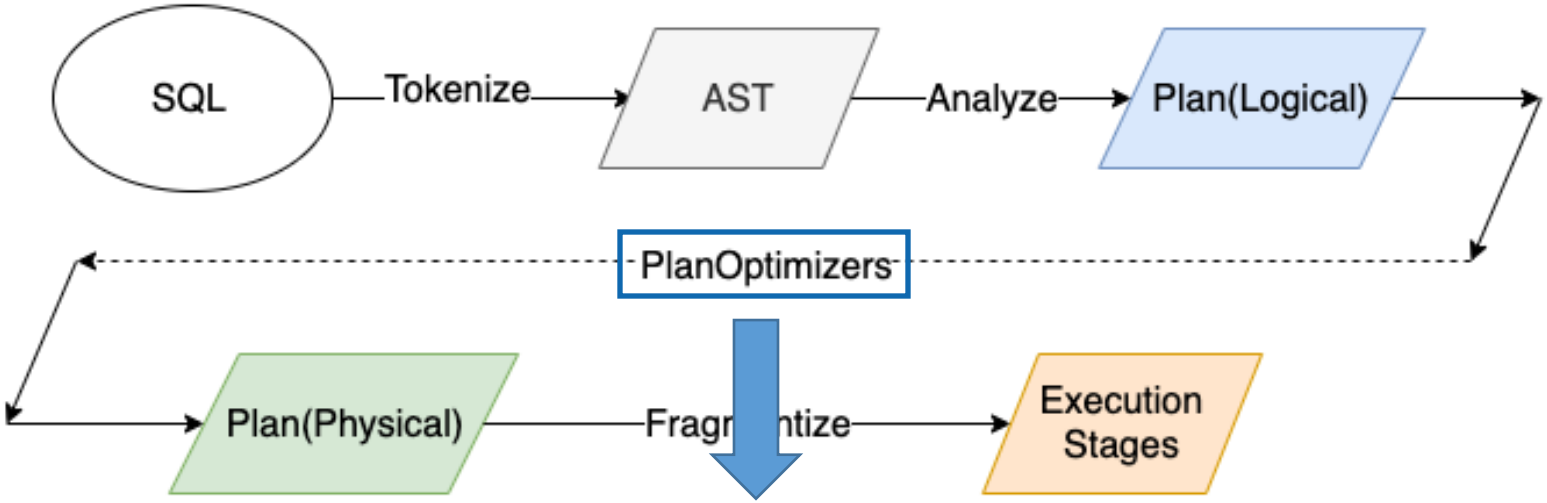
- 非下推情况
  - io.prestosql.plugin.jdbc.JdbcPlugin
  - io.prestosql.plugin.jdbc.JdbcClient
  - io.prestosql.plugin.jdbc.JdbcModule
  - io.prestosql.plugin.jdbc.BaseJdbcConfig
- 在openLooKeng kernel计算
- 无法利用数据源的优势



openLooKeng

https://openlookeng.io

openLooKeng 下推



TranslateExpressions

ApplyConnectorOptimization

Push down SQL

~~select * from tablename~~

openLooKeng

https://openlookeng.io

ClickHouse connector 实现

- 下推算子
  - 继承Base JDBC connector
  - ClickHouse非标准SQL
    - 对于表达式的微调整
    - 对函数的重写

```
@Override
public String aggregation(List<Selection> symbols, Optional<List<String>> groupingKeysOp, Optional<String> groupIdElementOP, String from)
{
    StringBuilder builder = new StringBuilder();
    builder.append(from);
    if (groupingKeysOp.isPresent()) {
        List<String> groupingKeys = groupingKeysOp.get();
        if (!groupingKeys.isEmpty()) {
            builder.append(" GROUP BY ");
            builder.append(Joiner.on(", ").join(groupingKeys));
        }
    }
    else if (groupIdElementOP.isPresent()) {
        String groupEleStr = groupIdElementOP.get();
        builder.append(" GROUP BY GROUPING SETS ");
        builder.append(groupEleStr);
    }
    return select(symbols, builder.toString());
}
```

openLooKeng

```
@Override
public String aggregation(List<Selection> symbols, Optional<List<String>> groupingKeysOp, Optional<String> groupIdElementOP, String from)
{
    StringBuilder builder = new StringBuilder();
    builder.append(from);
    if (groupingKeysOp.isPresent()) {
        List<String> groupingKeys = groupingKeysOp.get();
        if (!groupingKeys.isEmpty()) {
            builder.append(" GROUP BY ");
            builder.append(Joiner.on(", ").join(groupingKeys));
        }
    }
    else if (groupIdElementOP.isPresent()) {
        throw new UnsupportedOperationException("ClickHouse Connector does not support grouping");
    }
    return select(symbols, builder.toString());
}
```

ClickHouse

openLooKeng

https://openlookeng.io

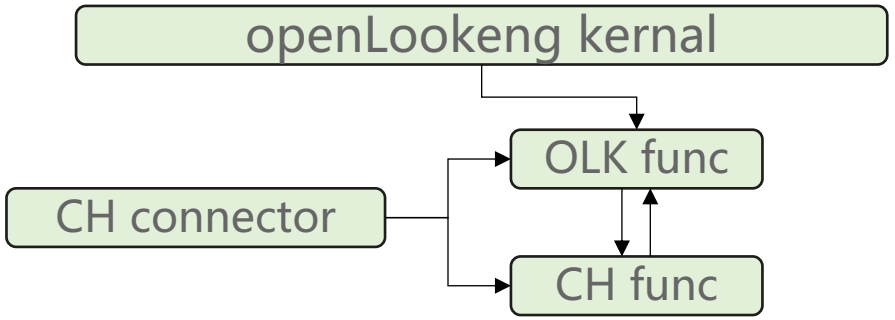
ClickHouse connector 实现

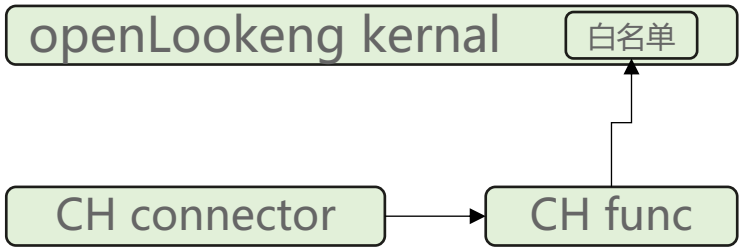
- 函数映射
  - 直接映射
    - 常用函数：聚合统计函数、数学函数、字符串函、时间日期函数
    - 不满足业务需求

函数类型	函数语法	适配开发处理
聚合统计函数	CORR($1,$2)，STDDEV($1)，stddev_pop($1)，stddev_samp($1)，skewness($1)，kurtosis($1)，VARIANCE($1)，var_samp($1)	大多数基本一致，直接映射即可
数学函数	ACOS($1)，ASIN($1)，ATAN($1)，ATAN2($1,$2)，CEIL($1)，CEILING($1)，COS($1)，e()，EXP($1)，FLOOR($1)，LN($1)，LOG10($1)，LOG2($1)，MOD($1,$2)，pi()，POW($1,$2)，POWER($1,$2)，RAND()，RANDOM()，ROUND($1)，ROUND($1,$2)，SIGN($1)，SIN($1)，SQRT($1)，TAN($1)	大多数基本一致，直接映射即可，部分函数需要对不同参数适配转换，分别有：1:n、n:1、n:m、1:m等类型。
字符串函数	CONCAT($1,$2)，LENGTH($1)，LOWER($1)，LTRIM($1)，REPLACE($1,$2)，REPLACE($1,$2,$3)，RTRIM($1)，STRPOS($1,$2)，SUBSTR($1,$2,$3)，POSITION($1,$2)，TRIM($1)，UPPER($1)	大多数基本一致，直接映射即可，部分函数需要对不同参数适配转换，分别有：1:n、n:1、n:m、1:m等类型。
时间日期函数	YEAR($1)，MONTH($1)，QUARTER($1)，WEEK($1)，DAY($1)，HOUR($1)，MINUTE($1)，SECOND($1)，DAY_OF_WEEK($1)，DAY_OF_MONTH($1)，DAY_OF_YEAR($1)	直接映射

ClickHouse connector 实现

- 函数映射
  - 外部函数的注册下推
    - ClickHouseExternalFunctionHub#getExternalFunctions
    - ClickHouseExternalDateTimeFunctions#getFunctionsInfo
  - 扩展实现
    - ApplyRemoteFunctionPushDown#rewriteRemoteFunction
    - BaseJdbcRowExpressionConverter#visitCall





openLooKeng

https://openlookeng.io

Connector结构

基本查询功能

openLooKeng函数下推

外部函数注册下推

openLooKeng

https://openlookeng.io

检查是否下推成功

```
explain select ckdb.function.toDateTime('123123123') from clickhouse152.ssb.customer limit 1
```

- explain

```
Output[_col0]
    Layout: [todatetime:timestamp]
    Estimates: {rows: ? (?), cpu: ?, memory: 0B, network: ?}
    _col0 := todatetime
  RemoteExchange[GATHER]
      Layout: [todatetime:timestamp]
      Estimates: {rows: ? (?), cpu: ?, memory: 0B, network: ?}
        TableScan[table = clickhouse152:clickhouse152.GeneratedSql{sql=SELECT toDateTime('123123123') AS todatetime FROM (SELECT null FROM ssb.customer LIMIT 1) hetu_table_1, isPushDown=true}]
        Layout: [todatetime:timestamp]
        Estimates: {rows: ? (?), cpu: ?, memory: 0B, network: 0B}
        todatetime := todatetime:timestamp:Optional[DateTime]
```

- 查询query log

```
SELECT
  event_time,
  query_duration_ms,
  read_rows,
  query
FROM
  system.query_log
where
  event_date = '2021-07-22'
  and type = 'QueryFinish'
  and query not like 'select timezone%'
  and query not like 'select version%'
  and query not like 'SELECT
  event_time,
  query_duration_ms,
  read_rows,
  query%'
  and query not like 'select database%'
  and query not like 'SELECT database%'
  and query not like 'select name as TABLE_SCHEM%';
```

openLooKeng

https://openlookeng.io

使用示例

- 配置文件：etc/catalog/clickhouseXXX.properties

```
connector.name=clickhouse
connection-url=jdbc:clickhouse://example.net:8123
connection-user=username
connection-password=yourpassword
```

- 允许连接器删除表

```
allow-drop-table=true
```

- 表名区分大小写

```
case-insensitive-name-matching=true
```

- 文档：https://openlookeng.io/zh-cn/docs/docs/connector/clickhouse.html

使用示例

- 配置外部函数命名空间

  - ckdb.properties

```
supported-function-languages=JDBC
function-namespace-manager.name=memory
```

  - clickhouseXXX.properties

```
connector.name=clickhouse
connection-url=jdbc:clickhouse:/[redacted]:8123
connection-user=[redacted]
connection-password=[redacted]
jdbc.pushdown.remotenamespace=ckdb.function
connector.externalfunction.namespace=ckdb.function
jdbc.pushdown-enabled=true
allow-drop-table=true
```

openLooKeng

https://openlookeng.io

查询演示

- SQL查询
- ckdb.function查询演示
- Show functions like ‘ckdb’
- 如何添加添加新的外部函数

openLooKeng

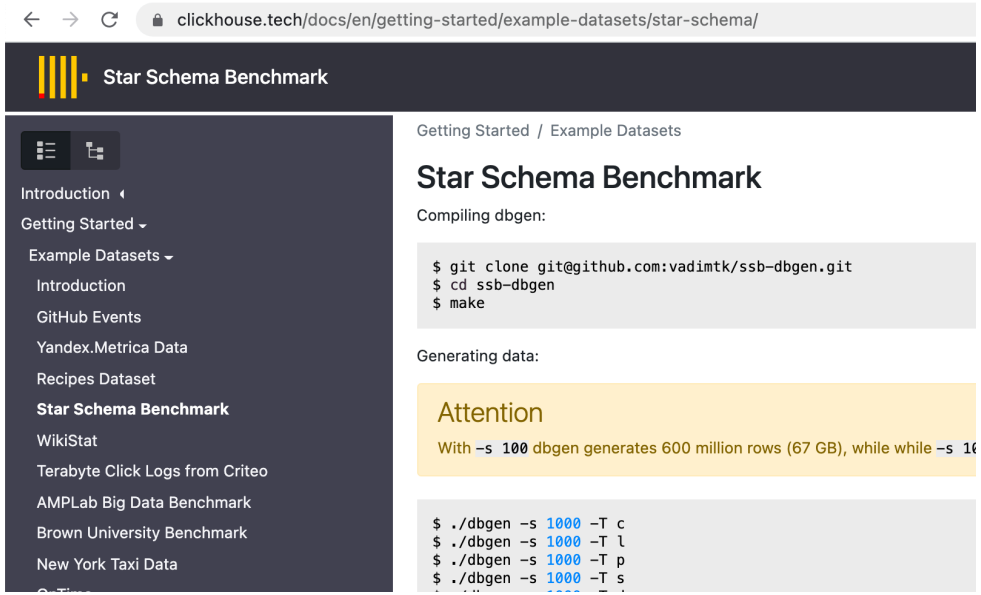
https://openlookeng.io

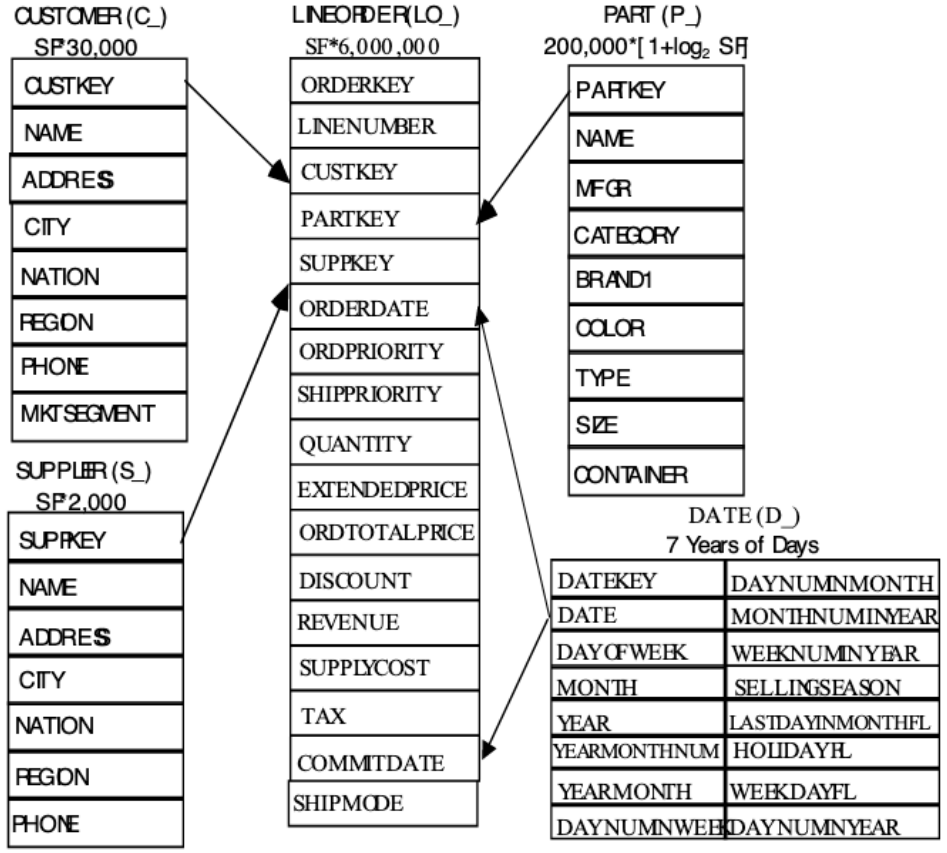
connector 目前的限制

- 类型支持暂不全面
  - 支持常见类型
  - Int128,Int256,UInt256,IPV4...
- 关于array的常用函数暂时不支持
- 不支持create，delete和alter操作
- 单表查询性能具有一定的损耗

CH connector性能测试

- **SSB**
  - Star Schema Benchmark
  - 基于TPC-H修改，并专门针对星型模型OLAP场景下的测试工具，共13条测试语句





openLooKeng

https://openlookeng.io

CH connector性能测试

- 使用ssb-dbgen工具构建了43亿条数据用于测试

table	size	bytes_on_disk	data_uncompressed_bytes	data_compressed_bytes	compress_rate	rows
customer	1.12 GiB	1.12 GiB	1.65 GiB	1.12 GiB	67.94630597	30000000
part	34.47 MiB	34.47 MiB	48.97 MiB	34.39 MiB	70.22835711	2000000
supplier	75.39 MiB	75.39 MiB	110.64 MiB	75.33 MiB	68.08099618	2000000
lineorder	123.00 GiB	123.00 GiB	176.19 GiB	122.76 GiB	69.67630518	4398761522
lineorder_flat	396.40 GiB	396.40 GiB	711.53 GiB	395.76 GiB	55.62024948	4398761522

CH connector性能测试

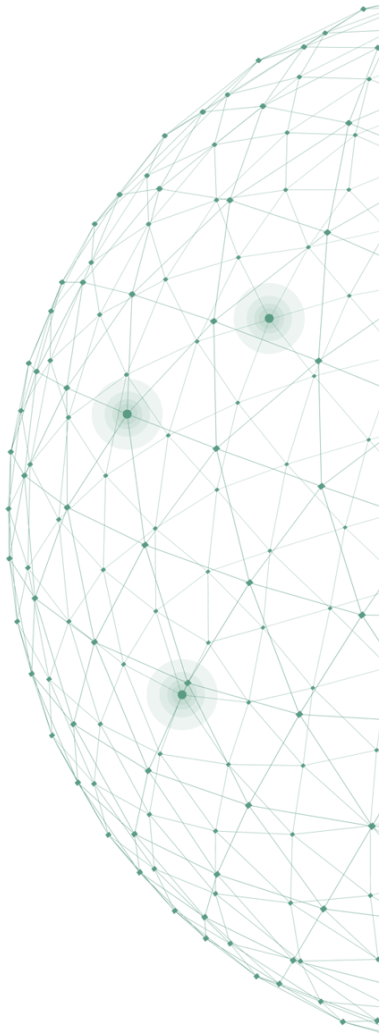
- 使用ssb-dbgen工具构建了43亿条数据用于测试

序号	Clickhouse	OpenLookeng	差异	百分比
Q1.1	14.07	16.10	-2.03	-14.41%
Q1.2	1.55	2.41	-0.87	-55.99%
Q1.3	0.50	0.67	-0.17	-34.00%
Q2.1	119.83	169.40	-49.57	-41.36%
Q2.2	5.53	9.93	-4.40	-79.53%
Q2.3	3.97	8.98	-5.01	-126.15%
Q3.1	39.90	68.20	-28.30	-70.93%
Q3.2	39.80	38.72	1.08	2.71%
Q3.3	12.28	20.89	-8.61	-70.11%
Q3.4	0.23	1.82	-1.59	-694.76%
Q4.1	107.18	191.40	-84.22	-78.57%
Q4.2	17.81	22.63	-4.82	-27.07%
Q4.3	11.40	19.39	-8.00	-70.17%
总计	**374.04**	**570.53**	**-196.49**	**-52.53%**

openLooKeng

https://openlookeng.io

Q&A



openLooKeng

https://openlookeng.io