Cloudera Streaming Analytics 1.3.0

SQL Stream Builder

Date published: 2019-12-17
Date modified: 2021-03-25

CLOUDERA

https://docs.cloudera.com/

Legal Notice

Contents

SQL Syntax Guide

The SQL Syntax Guide provides a collection of SQL syntaxes that is supported in SQL Stream Builder.

SQL Syntax

- Filtering WHERE
- Projections FOO || 'baz'
- Tumble, Hopping window expressions TUMBLE(), HOP()
- Grouping GROUP BY
- Filter after aggregation HAVING
- Join JOIN ON..
- Comparison operators
- Logical operators
- Arithmetic operators and functions
- Character string operators and functions
- Binary string operators and functions
- Date/Time functions
- Type conversion
- Aggregate functions
- Grouped window functions
- Grouped Auxiliary functions

Data types

- Datatypes

Metadata commands

- show tables - list virtual tables.
- desc <vtable> - describe the specified virtual table, showing columns and types.
- show jobs - list current running SQL jobs.
- show history - show SQL query history (only successfully parsed/executed).
- help - show help.

SQL Examples

You can use the SQL examples for frequently used functions, syntax and techniques in SQL Stream Builder (SSB). SSB uses Calcite Compatible SQL, but to include the functionality of Flink you need to customize certain SQL commands.

Metadata commands

```
-- show all tables
SHOW tables;
SHOW tables;

-- describe or show schema for table
DESCRIBE payments;
DESC payments;
```

Timestamps, intervals and time

```
-- eventTimestamp is the Kafka timestamp
-- as unix timestamp. Magically added to every schema.
SELECT max(eventTimestamp) FROM solar_inputs;

-- make it human readable
SELECT CAST(max(eventTimestamp) AS varchar) as TS FROM solar_inputs;

-- dete math with interval
SELECT * FROM payments
WHERE eventTimestamp > CURRENT_TIMESTAMP-interval '10' second;
```

Aggregation

```
-- hourly payment volume

SELECT SUM(CAST(amount AS numeric)) AS payment_volume,
CAST(TUMBLE_END(eventTimestamp, interval '1' hour) AS varchar) AS ts
FROM payments
GROUP BY TUMBLE(eventTimestamp, interval '1' hour);
```

```
-- detect multiple auths in a short window and
-- send to lock account topic/microservice

SELECT card,
MAX(amount) as theamount,
TUMBLE_END(eventTimestamp, interval '5' minute) as ts
FROM payments
WHERE lat IS NOT NULL
AND lon IS NOT NULL
GROUP BY card, TUMBLE(eventTimestamp, interval '5' minute)
HAVING COUNT(*) > 4 -- >4==fraud
```

Working with arrays

```
-- unnest each array element as separate row
SELECT b.*, u.*
 FROM bgp_avro b,
 UNNEST(b.path) AS u(pathitem)
```

Note: Arrays start at 1 not 0.

Union ALL

```
-- union two different tables
SELECT * FROM clickstream
WHERE useragent = 'Chrome/62.0.3202.84 Mobile Safari/537.36'
UNION ALL
SELECT * FROM clickstream
WHERE useragent = 'Version/4.0 Chrome/58.0.3029.83 Mobile Safari/537.36'
```

Math

```
-- simple math
SELECT 42+1 FROM mylogs;
```

```
-- inline
SELECT (amount+10)*upcharge AS total_amount
FROM payments
WHERE account_type = 'merchant'
```

```
-- convert C to F
SELECT (temp-32)/1.8 AS temp_fahrenheit
FROM reactor_core_sensors;
```

```
-- daily miles accumulator, 100:1
-- send to persistent storage microservice
-- for upsert of miles tally
SELECT card,
SUM(amount)/100 AS miles,
TUMBLE_END(eventTimestamp, interval '1' day)
FROM payments
GROUP BY card, TUMBLE(eventTimestamp, interval '1' day);
```

Joins

```
-- join multiple streams
select o.name,
       sum(d.clicks),
       hop_end(r.eventTimestamp, interval '20' second, interval '40' second)
 from click_stream o join orgs r on o.org_id = r.org_id
                join models d on d.org_id = r.org_id
 group by o.name,
        hop(r.eventTimestamp, interval '20' second, interval '40' second)
```

Hyperjoins

Joins are considered "hyperjoins" because SQLStreamBuilder has the ability to join multiple tables in a single query, and because a table is created from a data provider, these joins can span multiple clusters/connect strings, but also multiple types of sources (join Kafka and a database for instance).

```
SELECT us_west.user_score+ap_south.user_score
FROM kafka_in_zone_us_west us_west
FULL OUTER JOIN kafka_in_zone_ap_south ap_south
ON us_west.user_id = ap_south.user_id;
```

Misc SQL tricks

```
-- concatenation
SELECT 'testme_'||name FROM logs;
```

```
-- select the datatype of the field
SELECT eventTimestamp, TYPEOF(eventTimestamp) as mytype FROM airplanes;
```

Escaping and quoting

Typical escaping and quoting is supported.

- Nested columns

  ```
  SELECT foo.`bar` FROM table; -- must quote nested column
  ```

- Literals

```
SELECT "some string literal" FROM mytable; -- a literal
```

Built In Functions

```
-- convert EPOCH time to timestamp
select EPOCH_TO_TIMESTAMP(1593718981) from ev_sample_fraud;

-- convert EPOCH milliseconds to timestamp
select EPOCHMILLIS_TO_TIMESTAMP(1593718838150) from ev_sample_fraud;
```