Stream Analytics with SQL on Apache Flink®

Fabian Hueske
@fhueske

dataArtisans

Berlin Buzzwords
June, 13th 2017

Apache Flink

- Platform for scalable stream processing
- Fast
  - Low latency and high throughput
- Accurate
  - Stateful streaming processing in event time
  - Exactly-once state guarantees
- Reliable
  - Highly available cluster setup
  - Snapshot and restart applications

Powered by Flink

… and many more.

Flink's DataStream API

- The DataStream API is very expressive
  - Application logic implemented as user-defined functions
  - Windows, triggers, evictors, state, timers, async calls, …

- Many applications follow similar patterns
  - Do not require the expressiveness of the DataStream API
  - Can be specified more concisely and easily with a DSL

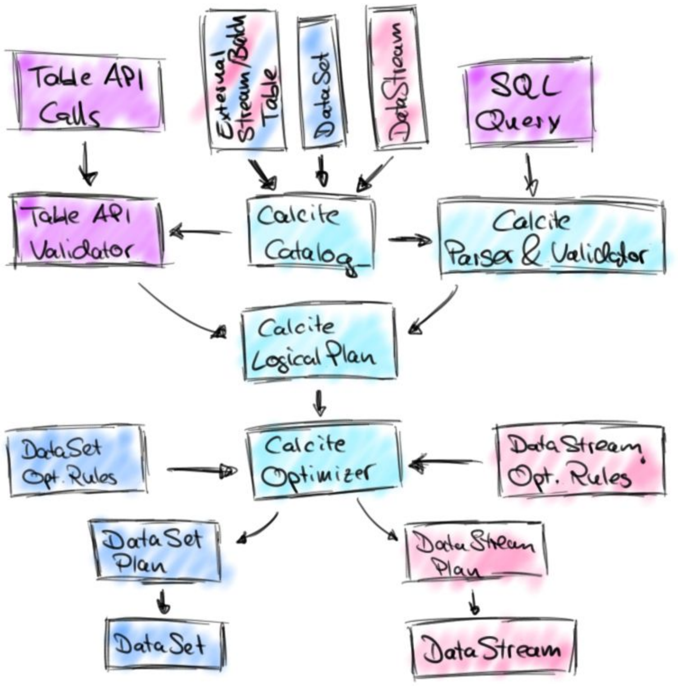
Q: What's the most popular DSL for data processing?
A: SQL!

Apache Flink's Relational APIs

- *Standard* SQL & LINQ-style Table API
- *Unified* APIs for batch & streaming data

A query specifies exactly the same result regardless whether its input is static batch data or streaming data.

- Common translation layers
  - Optimization based on Apache Calcite
  - Type system & code-generation
  - Table sources & sinks



Show me some code!

```
val tableApiResult: Table = tEnv
  .scan("clicks")
  .filter('url.like("https://www.xyz.com%")
  .groupBy('user)
  .select('user, 'link.count as 'cnt)
```

"clicks" can be a
- file
- database table,
- stream, …

```
val sqlResult: Table = tEnv.sql("""
  |SELECT user,
  |       COUNT(link) AS cnt
  |FROM clicks
  |WHERE url LIKE 'https://www.xyz.com%'
  |GROUP BY user
  """.stripMargin)
```

What if "clicks" is a file?

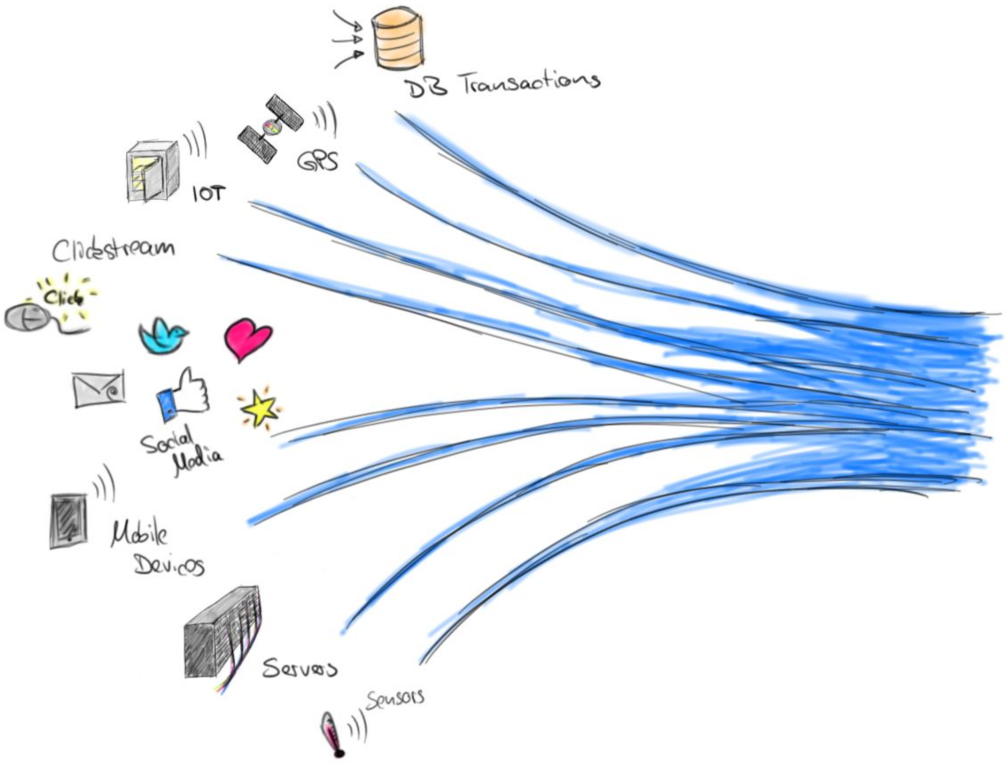
user	cTime	url	link
u1	12:00:00	https://…	l1
u2	12:00:00	https://…	l2
u1	12:00:02	https://…	l3
u3	12:00:03	https://…	l2

```sql
SELECT
  user,
  COUNT(link) as cnt
FROM clicks
GROUP BY user
```

user	cnt
u1	2
u2	1
u3	1

Q: What if we get more click data?
A: We run the query again.

What if "clicks" is a stream?



- We want the same results as for batch input!
- Does SQL work on streams as well?

SQL was not designed for streams

▪ Relations are bounded (multi-)sets.	↔	Streams are infinite sequences.
▪ DBMS can access all data.	↔	Streaming data arrives over time.
▪ SQL queries return a result and complete.	↔	Streaming queries continuously emit results and never complete.

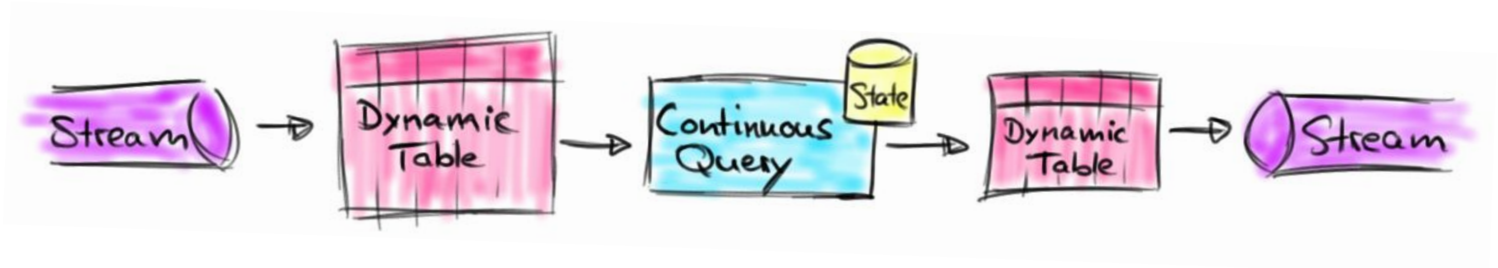
DBMSs run queries on streams

- Materialized views (MV) are similar to regular views, but persisted to disk or memory
  - Used to speed-up analytical queries
  - MVs need to be updated when the base tables change

- MV maintenance is very similar to SQL on streams
  - Base table updates are a stream of DML statements
  - MV definition query is evaluated on that stream
  - MV is query result and continuously updated

Continuous Queries in Flink

- Core concept is a *"Dynamic Table"*
  - Dynamic tables are changing over time
- Queries on dynamic tables
  - produce new dynamic tables (which are updated based on input)
  - do not terminate
- Stream ↔ Dynamic table conversions



Stream → Dynamic Table

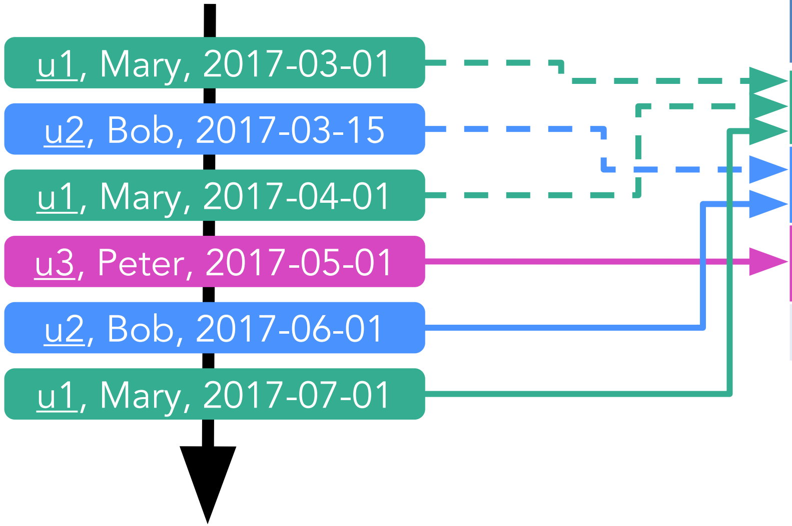
- Append mode
  - Stream records are appended to table
  - Table grows as more data arrives

u1, 12:00:00, https://…, l1
u2, 12:00:00, https://…, l2
u1, 12:00:05, https://…, l3
u3, 12:01:00, https://…, l2
u2, 12:01:30, https://…, l4
u1, 12:01:45, https://…, l2

user	cTime	url	link
u1	12:00:00	https://…	l1
u2	12:00:00	https://…	l2
u1	12:00:05	https://…	l3
u3	12:01:00	https://…	l2
u2	12:01:30	https://…	l4
u1	12:01:45	https://…	l2
…	…		

Stream → Dynamic Table

- Upsert mode
  - Stream records have (composite) key attributes
  - Records are inserted or update existing records with same key



user	name	lastLogin
u1	Mary	2017-07-01
u2	Bob	2017-06-01
u3	Peter	2017-05-01
…	…	

Querying a Dynamic Table

clicks

user	link
u1	l1
u2	l2
u1	l3
u3	l1
u3	l2
u1	l4

```
SELECT
  user,
  COUNT(link) as cnt
FROM clicks
GROUP BY user
```

result

user	cnt
u1	3
u2	1
u3	2

Rows of result table are updated.

What about windows?

```
val tableApiResult: Table = tEnv
  .scan("clicks")
  .window(Tumble over 1.hour on 'cTime as 'w)
  .groupBy('w, 'user)
  .select('user, 'w.end AS endT, 'link.count as 'cnt)


val sqlResult: Table = tEnv.sql("""
  |SELECT user,
  |       TUMBLE_END(cTime, INTERVAL '1' HOURS) AS endT,
  |       COUNT(link) AS cnt
  |FROM clicks
  |GROUP BY TUMBLE(cTime, INTERVAL '1' HOURS), user
  """.stripMargin)
```

Computing Window Aggregates

clicks		
user	**time**	**link**
u1	12:00:00	l1
u2	12:00:00	l2
u1	12:02:00	l2
u1	12:55:00	l4
u2	13:01:00	l1
u3	13:30:00	l4
u3	13:59:00	l3
u1	14:00:00	l1
u3	14:02:00	l2
u2	14:30:00	l2
u2	14:40:00	l4

```
SELECT
 user,
 TUMBLE_END(
   cTime,
   INTERVAL '1' HOURS)
  AS endT,
  COUNT(link) AS cnt
FROM clicks
GROUP BY
 user,
 TUMBLE(
  cTime,
  INTERVAL '1' HOURS)
```

result		
user	**endT**	**cnt**
u1	13:00:00	3
u2	13:00:00	1
u2	14:00:00	1
u3	14:00:00	2
u1	15:00:00	1
u2	15:00:00	2
u3	15:00:00	1

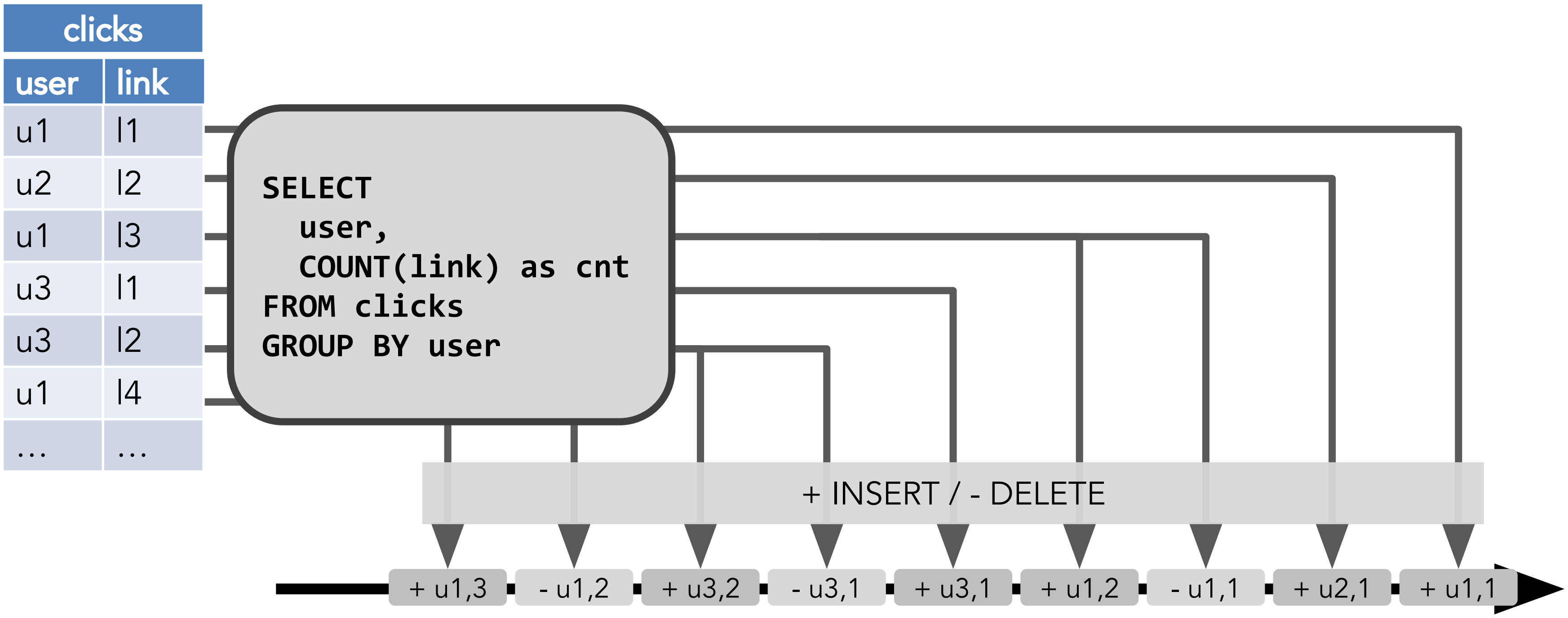
Rows are appended to result table.

Dynamic Table → Stream

- Converting a dynamic table into a stream
  - Dynamic tables might update or delete existing rows
  - Updates must be encoded in outgoing stream

- Conversion of tables to streams inspired by DBMS logs
  - DBMS use logs to restore databases (and tables)
  - REDO logs store new records to redo changes
  - UNDO logs store old records to undo changes

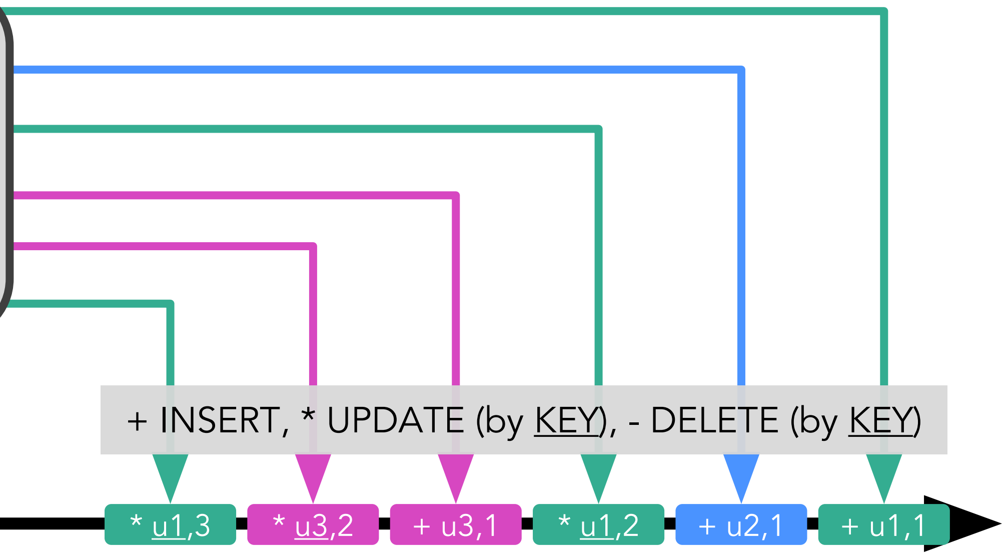
Dynamic Table → Stream: REDO/UNDO



Dynamic Table → Stream: REDO

clicks	
user	link
u1	l1
u2	l2
u1	l3
u3	l1
u3	l2
u1	l4
…	…

```
SELECT
  user,
  COUNT(link) as cnt
FROM clicks
GROUP BY user
```



Can we run any query on a dynamic table?

- No, there are space and computation constraints ☹
- State size may not grow infinitely as more data arrives

```
SELECT user, COUNT(link) FROM clicks GROUP BY user;
```

- A change of an input table may only trigger a partial re-computation of the result table

```
SELECT user, RANK() OVER (ORDER BY lastLogin) FROM users;
```

Bounding the Size of Query State

- Adapt the semantics of the query

```
SELECT user, COUNT(link) AS cnt
FROM clicks
WHERE last(cTime, INTERVAL '1' DAY)
GROUP BY user
```

  - Aggregate data of last 24 hours. Discard older data.

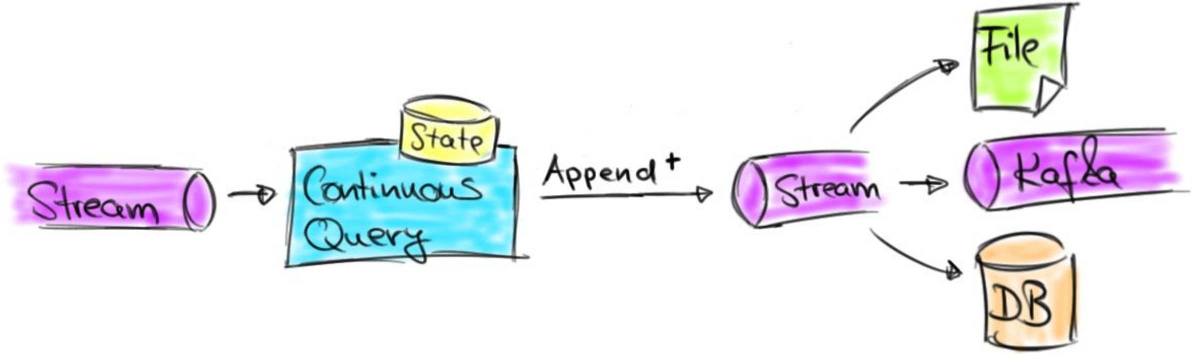
- Trade the accuracy of the result for size of state
  - Remove state for keys that became inactive.

Current State of SQL & Table API

- Flink's relational APIs are rapidly evolving
  - Lots of interest by community and many contributors
  - Used in production at large scale by Alibaba and others

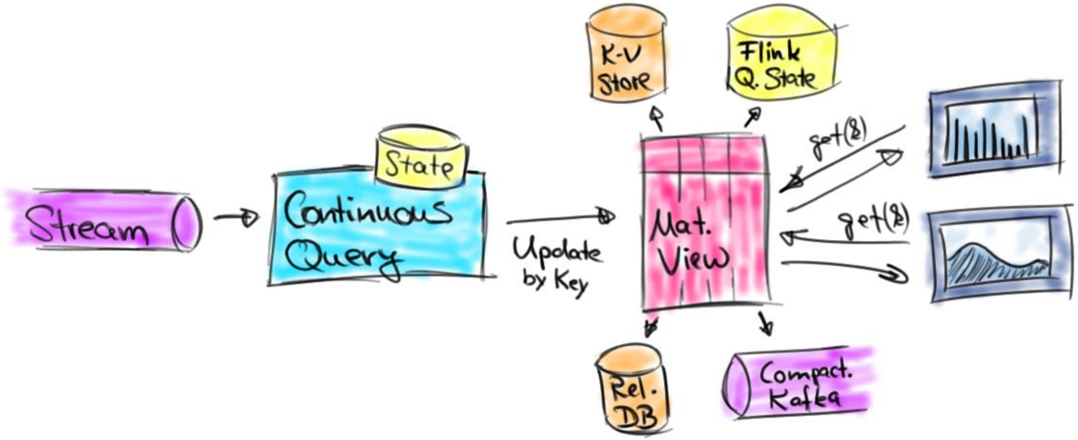
- Features released in Flink 1.3.0
  - GroupBy & Over windowed aggregates
  - Non-windowed aggregates (with update changes)
  - User-defined aggregation functions

What can be built with this?



- Continuous ETL
  - Continuously ingest data
  - Process with transformations & window aggregates
  - Write to files (Parquet, ORC), Kafka, PostgreSQL, HBase, …

What can be built with this?



- Dashboards, reporting & event-driven architectures
  - Flink updates query results with low latency
  - Result is written to KV store, DBMS, compacted Kafka topic
- Later, results can be maintained as queryable state

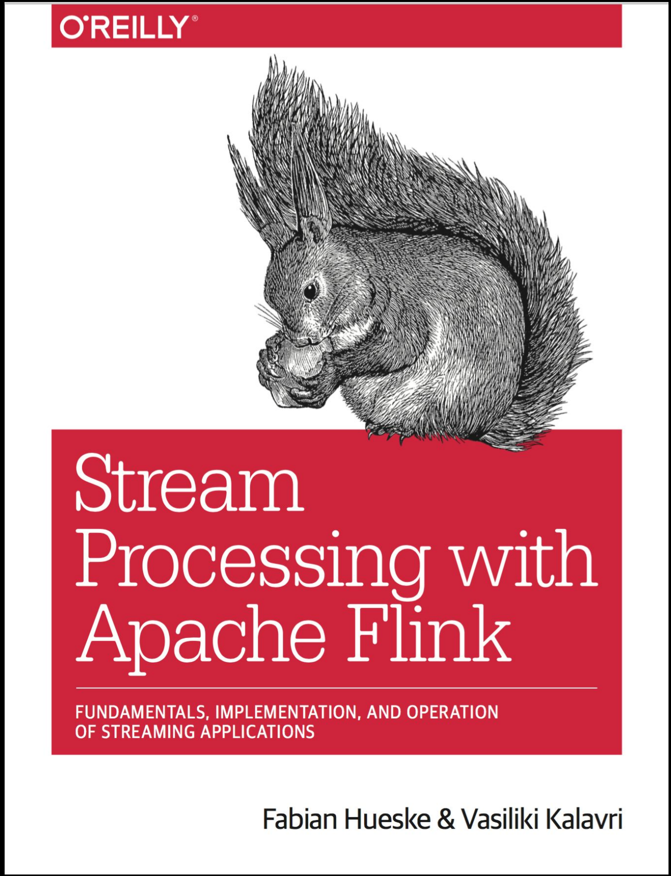
Conclusion

- Table API & SQL support many streaming use cases
  - High-level / declarative specification
  - Automatic optimization and translation
  - Efficient execution
  - Scalar, table, aggregation UDFs for flexibility

- Updating results enable many exciting applications

- Check it out!


FLINK
FORWARD
Berlin
11-13 Sep 2017
Flink Forward, the premier conference on Apache Flink®, is coming back to Berlin
Call for Submissions is open



Thank you!

@fhueske
@ApacheFlink
@dataArtisans

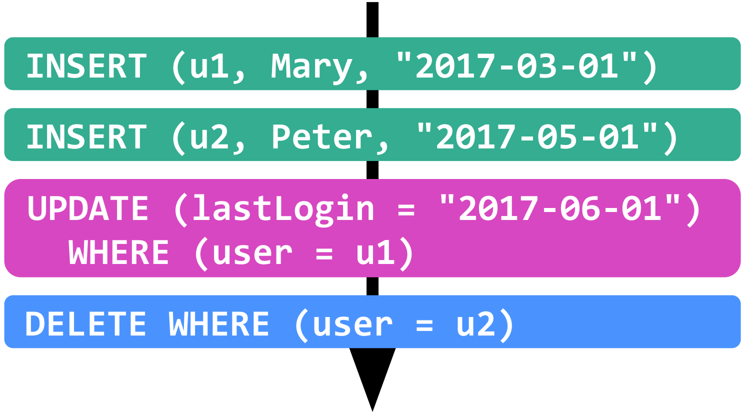
Available on O'Reilly Early Release!

We are hiring!

data-artisans.com/careers

Tables are materialized streams

- A table is the materialization of a stream of modifications
  - SQL DML statements: INSERT, UPDATE, and DELETE
  - DBMSs process statements by modifying tables



user	name	lastLogin
u1	Mary	2017-06-01
u2	Peter	2017-05-01

About me

- Apache Flink PMC member
    - Contributing since day 1 at TU Berlin
    - Focusing on Flink's relational APIs since 1.5 years

- Co-author of "Stream Processing with Apache Flink"
    - Work in progress…

- Co-founder of data Artisans

dataArtisans

dA

PLATFORM

Original creators of **Apache Flink®**

Providers of the **dA Platform**, a supported Flink distribution