

dataArtisans



Apache Flink® Training

Table API

June 15th, 2015

Agenda



- Table API Overview
- Expressions
- DataSet to Table
- Table to DataSet

Table API

Table API Overview



- Makes analysis of structured data very easy
 - Think of database tables (CREATE TABLE)
- Evaluates SQL-like expressions
 - Code generation for execution
- Tight integration with DataSet API
 - Convert DataSet to Table and back

Table API Overview



- Basic data structure is a Table
- Table is structured data with named fields
 - Similar to relational table
- Expressions evaluated on a Table yield a new Table

Table API Dependency



- In order to use the Table API in your program, add this to the `<dependencies>` section of your `pom.xml`

```
<dependency>
  <groupId>org.apache.flink</groupId>
  <artifactId>flink-table</artifactId>
  <version>0.9-SNAPSHOT</version>
</dependency>
```

Table API Expressions



- `Table t = orig.as("author, title, pages");`
`// filter table`
- `Table t2 = t.filter("pages > 100");`
`// project table`
- `Table t3 = t.select("author, title");`
- `Table t4 = t.select("pages*2 as dPages");`
`// group table and compute aggregations`
- `Table t5 = t.groupBy("author")`
`.select("pages.avg as avgPages");`
`// join two table`
- `Table t6 = t.join(t.select("author2, title2"))`
`.where("author = author2")`
`.select("title, title2");`

DataSet to Table



Java DataSet API via TableEnvironment:

```
// data set
DataSet<Table3<String,Long,Double>> ds = ...;

// get a TableEnvironment
TableEnvironment tEnv = new TableEnvironment();

// convert data set to Table and give name to fields
Table t = tEnv.fromDataSet(ds)
               .as("name, count, price");
```


Table to DataSet



- Java DataSet API via TableEnvironment
- Convert to custom POJO data set
 - Pojo fields must map to Table fields

```
public static class Stock {  
    public String name;  
    public int count;  
    public double price;  
}
```

```
Table t = x.as("name, count, price");
```

```
TableEnvironment tEnv = new TableEnvironment();
```

```
DataSet<Stock> ds = tEnv.toDataSet(t, Stock.class);
```

Table to DataSet



Java DataSet API via TableEnvironment

- Convert to DataSet<Row>
- Most valuable for printing

```
Table t = x.as("name, count, price");  
TableEnvironment tEnv = new TableEnvironment();  
tEnv.toDataSet(t, Row.class).print();
```

