

Efficiently Handling Streams from Millions of Sensors

Wednesday, October 11, 2017 4:00 PM (30 minutes)

We present two research works dealing with massive sensor data inputs.

1) We present I², an interactive development environment for real-time analysis pipelines, which is based on Apache Flink and Apache Zeppelin. The sheer amount of available streaming data frequently makes it impossible to visualize all data points at the same time. I² coordinates running cluster applications and corresponding visualizations such that only the currently depicted data points are processed in Flink and transferred towards the front end. We show how Flink jobs can adapt to changed visualization properties at runtime to allow interactive data exploration on high bandwidth data streams. Moreover, we present a data reduction technique which minimizes data transfer while providing loss free time-series plots.

2) We present Cutty, an innovative technique for the efficient aggregation of user-defined windows over data streams. While the aggregation of periodic sliding and tumbling windows was extensively studied in the past, little to no work was done on optimizing the aggregation of common, non-periodic windows. Typical examples of non-periodic windows are punctuation windows and sessions which can implement complex business logic. Cutty performs aggregate sharing for data stream windows, which are declared as user-defined functions (UDFs) and can contain arbitrary business logic. Cutty outperforms the state of the art for aggregate sharing on single and multiple queries. Moreover, it enables aggregate sharing for a broad class of non-periodic UDWs.

We close the talk with an outlook on the ongoing research of the Berlin Big Data Center regarding the efficient processing of data from millions of sensors.

Track

BDAHMM

Author: Mr TRAUB, Jonas (Technische Universität Berlin)

Presenter: Mr TRAUB, Jonas (Technische Universität Berlin)

Session Classification: Analytics