HStreaming: In-Memory Analytics of Petabytes of Data

Jan Stoess

The majority of Big Data is machine generated such as operational data, time series data, and video feeds. While Hadoop is an excellent tool for deriving insights across huge data sets, it is too slow for real-time analytics which requires immediate automated action. In this talk, Jan Stoess will discuss architectural approaches and challenges to real-time big data analytics in production deployments. He will cover the full data pipeline from real-time data acquisition, real-time processing, continuous archival, batch analytics, and visualization.