*Abstract*—We introduce Cloud DIKW as an analysis environment supporting scientific discovery through integrated parallel batch and streaming processing. This builds on common Big Data tools including Storm, ActiveMQ, Hbase and Hadoop and is tested on a social media data stream clustering application. Here the data points (such as tweets) are represented as high-dimensional vectors that reflect both textual content and social network information. This enables high quality clustering but due to the high cost of similarity computation, sequential implementations of even single-pass algorithms cannot keep up with the speed of real-world streams. This application is used to test the effectiveness of Cloud DIKW to support parallel clustering and meet real time constraints. This paper focusses on two systems-level issues. Firstly, most stream processing engines such as Apache Storm, organize distributed workers in the form of a directed acyclic graph (DAG), which makes it difficult to dynamically synchronize the state of parallel clustering workers. We tackle this challenge by creating a separate synchronization channel using a pub-sub messaging system (ActiveMQ in our case). Secondly, due to the sparsity of the high-dimensional vectors, the size of centroids grows quickly as new data points are assigned to the clusters. As a result, traditional parallel clustering synchronization that directly broadcasts centroids, becomes too expensive and limits the scalability of the parallel algorithm. In this paper, we address this problem by communicating only dynamic changes of the clusters rather than the whole centroid vector. Our algorithm under Cloud DIKW can process 10% of Twitter data in real-time with 96-way parallelism. By natural improvements to Cloud DIKW, including advanced collective communication techniques developed in our Harp project, we will be able to process the full Twitter data in real-time with 1000-way parallelism. Further our use of powerful general software subsystems will enable many other applications that need integration of streaming and batch data analytics.