CASE STUDY

STREAMANALYTIX
An Impetus Product

Real-time Multi-lingual Classification and Sentiment Analysis of Text

Highlights and Benefits

- *Rapid and accurate real-time text categorization and sentiment analysis*
- *Adjustable text categorization for domain-specific classes*
- *Multi-lingual support*
- *Enhanced sentiment analysis to focus on feature-specific opinion mining*
- *Linear scalability to increase the number of nodes in the cluster*
- *Provision to add custom component for added functionalities*

Client Overview

The client is a major telecom company providing nationwide telecom services. They wanted a system that performs real-time, multi-lingual classification and sentiment analysis of text data.

Challenges

The client was looking for a solution that allows storing, indexing, and querying PetaBytes (PBs) of data with a very high throughput. Some of the critical requirements were:

- Ingest and parse high volume of data [250M (15 TB) records/day] of varied types (for example, weblogs, email, chat, and files)
- Apply real-time multi-lingual classification and sentiment analysis with very high accuracy (four nines)
- Store metadata and raw binary data for querying
- Query SLA - 5s on cold data

Technologies

R, Latest Semantic Analysis, Text Mining, Apache Kafka, Apache HBase, Elasticsearch, Apache Storm

Our Solution

The solution provided by Impetus had three modules:

- Analytics Module: Responsible for performing text categorization and sentiment analysis. It implements a matrix decomposition-based text-classification algorithm. The incoming test document had to pass through a series of pre-processing and numerical computations. Impetus designed the classifier to achieve very low latency.
- Event Store/ Indexer Abstraction Layer: Responsible for storing and indexing the information based on the configuration
- Publish Module: Responsible for publishing the analytical result or event data to the external system

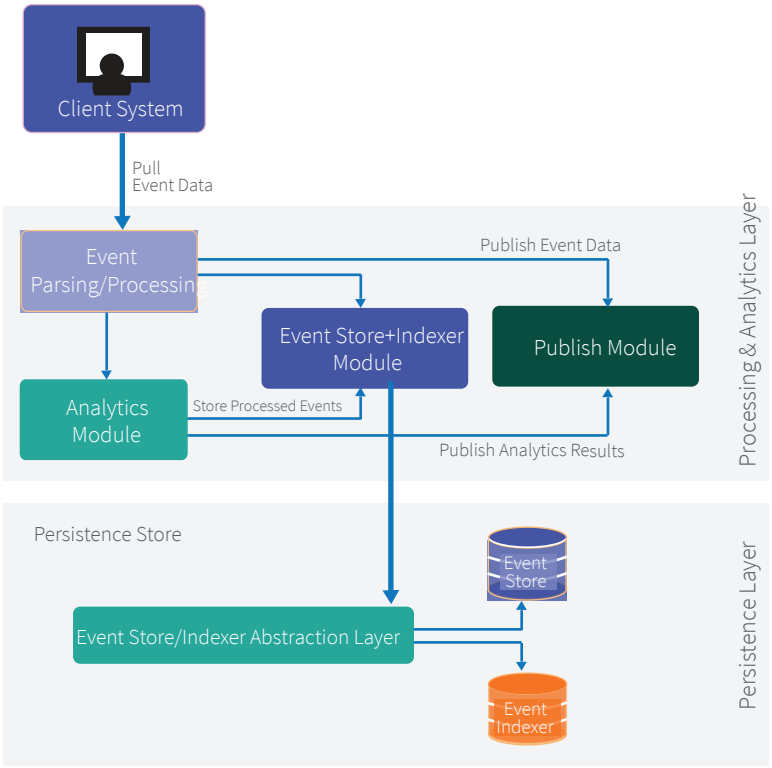

Figure 1: Solution Architecture

STREAMANALYTIX
An Impetus Product

StreamAnalytix is an enterprise grade, visual, big data analytics platform for unified streaming and batch data processing based on best-of-breed open source technologies. It supports the end-to-end functionality of data ingestion, enrichment, machine learning, action triggers, and visualization, StreamAnalytix offers an intuitive drag-and-drop visual interface to build and operationalize big data applications five to ten times faster, across industries, data formats, and use cases.

Visit **www.streamanalytix.com** or write to us at **inquiry@streamanalytix.com**