IoT and Big Data platform

In IoT where sensors are connected in a sensor network, some of the processing may be done at edge of the network. For example, there may sensor hubs within the sensor network. Each hub is connected to a set of sensors, which may do some processing for the sensors it’s connected to.

However, most data will still be funneled into a Big Data platform for real time analytic.

Why Realtime Time Series Analytic

Big Data typically handles data of different kind. In the connected everything or Internet of Things world, machine / device generated data is mostly time series data. Typically insight from such data needs to be learned in real time.

Ruscello

*Ruscello* provides a real-time time series analytic platform particularly with IoT in mind. Although the different use cases are implemented in Spark Streaming, it is possible to implement them in Storm. The current focus is Spark Streaming implementation

The core time series analytic algorithms are implemented as Java API in another project, which enables the implementation of use cases in any real time clustering framework, whether spark-streaming or Storm.

Configuration

*Ruscello* believes in giving power in the hands of the end users. There are many configuration parameters to control the behavior of the solution.

By focusing on application specific configuration, the underlying engine remains as generic as possible.

Features

Ruscello is new project, but evolving very fast. Analytic is performed on windowed data. The following features are available now. More will be coming

* Basis statistics (mean, standard deviation etc)
* Time series outlier (e.g. abnormal ECG signal)
* Time series similarity (pattern detection)
* Time series level shift detector (some sensor that is expected to maintain a steady reading)

Example Use Case

This a specific use case for time series outlier detection. The specificity is obtained purely by configuration.

A transportation company ships pharmaceutical products inside container that is supposed to maintain certain temperature range. The pharmaceutical company has an SLA with the transportation company. If the SLA is broken, the transportation company is liable for damage.

The real time temperature sensor data is streamed to spark streaming, where analytic is performed to detect violation of SLA.

The raw data and the SLA breach data could be saved in a NOSQL database like Cassandra.