WHITE PAPER

Scalable real time data exploration, analysis and anomaly detection for IoT solutions

An IoT solution monitors large volumes of IoT devices in real time, processing millions of messages along with real time data exploration, analysis and anomaly detection all at once. Connected devices need to be monitored for anomalies so that a central operator can take immediate actions and resolve them quickly.

Azure® and AWS®, and out-of-the-box, fully managed services like Amazon SageMaker® and Azure Time Series Insights offer anomaly detection methods of sensor data to conduct root-cause analyses, and avoid costly downtime of IoT devices.

THINXTREAM®

Introduction

IoT solutions enables millions of varying types of globally distributed physical devices and sensors to be connected over cloud and exchange data between them.

Connected devices need to be monitored for anomalies so that a central operator can take immediate actions and resolve them quickly. This requires large volumes of dissimilar telemetry data from devices to be analyzed by a cloud server, which then needs to respond to identified anomalies in real time. However, monitoring and gathering insights from massive scale of IoT data is error-prone and often not fully automated , making it hard to scale effectively.

Scalable real time data exploration, analysis and anomalies in historical and real time sensor data can be detected by applying machine learning.

In this white paper, we examine the application of Azure and AWS machine learning modules, and an out-of-the-box Azure service Time Series Insights to detect anomalies in sensor data, conduct root-cause analyses, and avoid costly downtime of IoT devices.

Azure Machine Learning

Anomalies in device telemetry data can be sensed by Azure built-in or custom machine learning modules. Azure provides an Anomaly Detection module as part of Azure Machine Learning Studio and Anomaly Detection API. It can detect spikes and dips, positive and negative trends, level changes and changes in dynamic range of values in the time series data. They are customizable and can be trained according to the needs.

The scores generated by the Anomaly Detection API provide insights for generating alerts which is monitored through dashboards or connected with other business intelligence systems. Power BI® provides a rich dashboard for real time data and predictive analytics visualizations.

The Anomaly Detection API is effective and efficient for anomaly detection. Also at streaming level, the Azure “AnomalyDetection” operator is used to detect sustained increase or decrease in the level of values, slow negative and positive trends in the event stream.

Azure Machine Learning Studio provides a visual drag and drop workspace to easily build and test machine learning models. These models can be published from Machine Learning Studio as Web services that can be consumed by other external applications. The required datasets from all sources are added to an experiment and then cleaned, transformed and analyzed through various

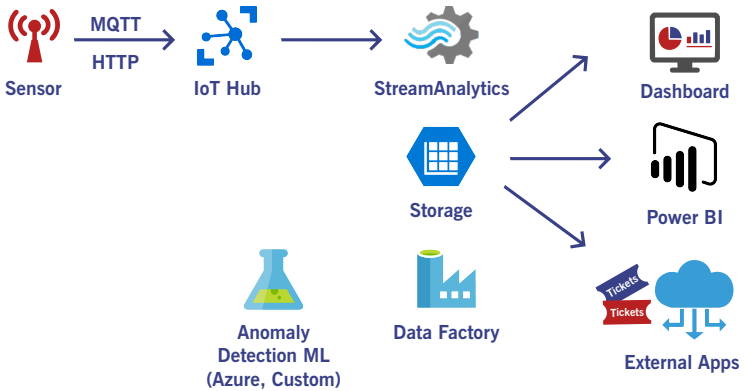

Figure 1. A typical Azure IoT Hub reference architecture of anomaly detection with Machine Learning

data manipulation and statistical functions. Cortana® Intelligence Suite, a fully managed big data and advanced analytics suite can also be used to transform the data. Finally a trained predictive analysis model is built from the algorithm modules.

In a typical IoT solution, Web Services implementing Machine Learning models built using Azure Machine Learning Studio would be integrated with Azure IoT Hub and other Azure Services for analyzing real time data. The sensor data received by the Azure IoT Hub is analyzed by Azure Stream Analytics, and further streamed out and logged in Azure Storage, and then fed into a Data Factory pipeline, which does the analysis using the Anomaly Detection Machine Learning Algorithms after the required cleaning and data transformation. The anomalies are visualized in dashboard user interface or are consumed by other Business Intelligence Systems workflows.

Amazon SageMaker

Figure 2. A typical anomaly detection workflow with Amazon SageMaker

The Amazon SageMaker is a fully managed solution for training and deploying machine learning models by using various built in algorithm modules. The Amazon SageMaker Random Cut Forest (RCF) is an unsupervised learning algorithm which detects anomalous data points and associates a score with each data point. Highest anomaly scores of more than three standard deviation from the mean score indicates likely occurrence of an anomaly.

In a systematic and industry standard approach to leverage SageMaker's Anomaly Detection capabilities, the first step is to preprocess and clean the data by removing duplicate and unwanted data to improve accuracy. A commonly used preprocessing strategy is by shingling where the data is converted into s-dimensional vectors to filter out the noises. Further, the algorithm is run by specifying the training job parameters against large volume of accurate and disparate data to finally develop a machine learning model, which is used for anomaly detection with real time streaming data.

Time Series Insights

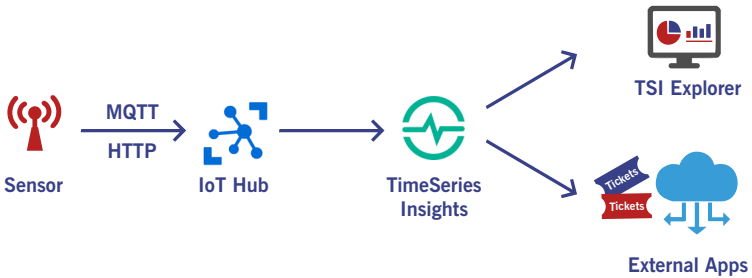

Figure 3. A simple reference architecture of anomaly detection with Time Series Insights

Times Series Insights (TSI) is a fully managed Azure service for near real time data exploration, storing, querying, analysis and visualization of time series data in the cloud. TSI connects to Azure IoT Hub or Event Hub event sources to collect the data and convert the telemetry JSON data to columnar data with indices. This cleaned data is put into its managed storage for easy retrieval and retained in persistent storage. The data can be sliced and diced easily using the out-of-the-box TSI Explorer and can be viewed in an intuitive and very simple way.

TSI can directly connect to multiple Azure IoT Hub or Event Hubs and anomalies can be spotted in TSI Explorer or it can be fed into other external applications via TSI REST APIs so that the TSI backend does the indexing, storing and aggregating of the time series data.

TSI can operate without any known schema and can identify the changes in overall trend and changes in range of values. TSI offers capabilities like charts, heat maps, and data drill downs to spot anomalies and trends quickly and in real time

TSI provides enterprise level security by using Azure security services like Azure Active directory (AAD), Single Sign On (SSA), Multi Factor Authentication (MFA) and Role Based Access Control (RBAC). TSI can also provide a global view of time series data across geographies in a global scale.

Conclusion

Connected devices need to be monitored for anomalies so that a central operator can take immediate actions and resolve them quickly. Monitoring and gathering insights from massive scale of IoT data is often not fully automated and hence error-prone, making it hard to scale effectively.

Scalable real time data exploration, analysis and anomalies in historical and real time sensor data can be detected by applying machine learning. Azure Machine Learning, Amazon SageMaker and Azure TSI provide tools to build services that can be integrated into various business workflows to provide better visibility into the massive scale of IoT devices across multiple locations in a secure manner.

As an IoT services provider, Thinxtream has expertise in machine learning for IoT solutions across Azure IoT Hub and AWS IoT. By leveraging the IoT expertise built over a decade, Thinxtream ensures cost-effective, quality and timely delivery of IoT solutions.

References

- https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/apps-anomaly-detection-api
- https://aws.amazon.com/blogs/machine-learning/use-the-built-in-amazon-sagemaker-random-cut-forest-algorithm-for-anomaly-detection/
- https://docs.aws.amazon.com/sagemaker/latest/dg/randomcutforest.html
- https://docs.microsoft.com/en-in/azure/time-series-insights/

Thinxtream Technologies is a global software company with a portfolio of innovative software platforms, products, components, solutions, patents, competences and services for Internet of Things (IoT) across several industry verticals and applications, successfully enabling leading customers, including Fortune 500 companies, meet their application, product and business goals.

Interested in learning more? For more information contact:

Thinxtream Technologies Pte. Ltd.
220 Orchard Road #05-01
Midpoint Orchard
SINGAPORE 238852
Phone: +65 66358625
Email: info@thinxtream.com

Thinxtream Technologies, Inc.
10260 SW Greenburg Road
Suite 400 Portland, OR 97223,
U.S.A
Phone: +1 503 293-3598
Email: info@thinxtream.com

www.thinxtream.com

LinkedIn/thinxtream

TT-WP-008-1-0918

THINXTREAM®