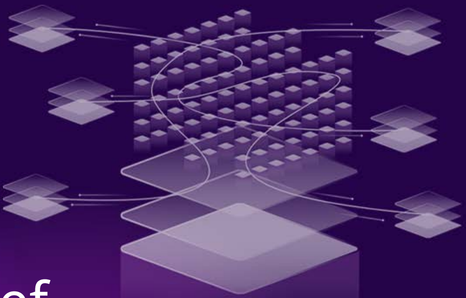


AIoT Solution Brief

Empowering IoT with AWS Greengrass and Edge Computing

Wilhelm Almer (VividCloud)

Aug 29, 2024

(781) 654-7800
www.vividcloud.com

HEADQUARTERS
150 Admiral Fitch Ave
Brunswick, ME 04011

NEW HAMPSHIRE
100 Domain Drive
Exeter, NH 03833

MASSACHUSETTS
85 Swanson St.
Boxborough, MA

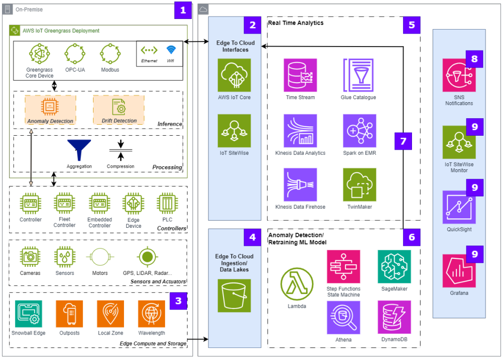
AIoT Solution Overview

- This solution offering leverages AWS services to provide an end-to-end AIoT platform designed to optimize industrial operations through real-time data ingestion, anomaly detection, and machine learning.
- AIoT, or Artificial Intelligence of Things, combines AI and the Internet of Things to improve IoT-related operations and provide gains in data analysis.
- This platform integrates edge computing, cloud processing, and digital twin technologies to offer comprehensive visibility and control over industrial assets, enhancing operational efficiency and reducing downtime.

Use Cases

- **Predictive Maintenance:** Detect and address equipment failures before they occur by leveraging real-time anomaly detection and machine learning.
- **Operational Efficiency:** Optimize production lines and reduce waste by analyzing telemetry data and using digital twins to simulate and improve processes.
- **Compliance and Reporting:** Maintain detailed logs of machine performance and anomaly detection for regulatory compliance and operational reporting.

AIoT Reference Design



1 Telemetry from industrial assets like PLCs, sensors and more is ingested by AWS IoT Greengrass leveraging pre-fabricated connectors.

Hot anomaly detection originates at the edge, performing regular data stream analytics and machine learning inference.

2 Edge to Cloud ingestion can be performed in real-time when network connectivity is available. AWS IoT Core ingests data via an MQTT interface for further processing. IoT SiteWise provides aggregation, asset and model management, specifically geared towards IoT assemblies.

Scheduled AWS IoT Jobs perform detection on the edge to identify machine learning (ML) result drift.

3 Edge and close-to-edge compute and storage is used to store raw device telemetry as well as ML inference results - but it is also used to retrain models and perform data analysis.

Terabytes of data can be ingested into the data lake more effectively to perform anomaly detection in the Cloud and to retrain existing (or train new) ML models.

4 An AWS S3 data lake architecture stores raw and processed telemetry data, trained machine learning (ML) models and ML inference results.

5 Amazon Kinesis Data Analytics runs queries to determine anomalistic behavior in datasets. Kinesis Data Analytics can be used to filter or process telemetric data before storing time series records.

Time Stream database stores time series data including chosen dimensions.

6 Perform cold anomaly detection leveraging cold data in the AWS S3 data lake in AWS S3. After drifts are detected - in the cloud, or at the edge based on hot data - Lambdas and step functions play key roles to orchestrate ML training and analysis.

7 Push re-trained ML model to the IoT device, to perform hot data anomaly detection.

8 Operational technology teams consume alerts from SNS as emails, text messages, or integration into ticketing systems.

9 No-code dashboards assess real-time and historical machine performance.

Telemetry Ingestion from Industrial Assets

- **Sources:** The workflow begins with telemetry data being generated from various industrial assets, including Cameras, PLCs (Programmable Logic Controllers), sensors (such as environmental sensors), and other edge devices.
- **Edge Processing:** AWS IoT Greengrass is deployed on the edge devices to process this data locally. The Greengrass deployment utilizes pre-fabricated connectors to ingest telemetry data efficiently. This local processing enables hot anomaly detection by performing real-time data stream analytics and machine learning inference directly at the edge.

Edge-to-Cloud Data Ingestion

- **AWS IoT Core:** When network connectivity is available, the processed data is ingested into the cloud via AWS IoT Core using MQTT interfaces. This service acts as the primary gateway for IoT devices to communicate with cloud services.
- **AWS IoT SiteWise:** For more complex IoT assemblies, AWS IoT SiteWise provides additional capabilities for aggregation, asset, and model management. SiteWise facilitates the collection, storage, and visualization of industrial equipment data.

Cloud Data Storage and Anomaly Detection

- **Data Lake (AWS S3):** The ingested telemetry data, along with machine learning models and inference results, are stored in a data lake architecture built on AWS S3. This architecture supports the storage of both raw and processed data.
- **Cold Anomaly Detection:** Cold data stored in the data lake is analyzed for anomalies. AWS Kinesis Data Analytics runs queries on the data to detect any abnormal behavior in the datasets. Time series data, including various dimensions, is stored in AWS Timestream, which can be filtered or processed further by Kinesis before storing records.

Machine Learning Model Training and Deployment

- **AWS SageMaker:** Machine learning models are trained and retrained in the cloud using AWS SageMaker. These models are then pushed back to the edge devices to enhance the capability of real-time anomaly detection (hot data) at the edge.
- **Drift Detection and Retraining:** The system continuously monitors for drift in machine learning results, both at the edge and in the cloud. AWS Step Functions and Lambda are used to orchestrate the retraining and deployment of ML models when drift is detected.

Monitoring and Visualization

- **AWS Grafana and QuickSight:** Real-time and historical performance of machines and assets are monitored and visualized using no-code dashboards provided by AWS Grafana and QuickSight. These tools enable operational technology (OT) teams to assess system performance at a glance.
- **Alerts and Notifications:** Alerts generated from anomaly detection and other monitoring activities are pushed to operational teams through AWS SNS, which can send notifications via email, SMS, or integrate with ticketing systems.

Edge and Close-to-Edge Computing

- **Local Processing:** In scenarios where large amounts of data are generated, edge and close-to-edge compute and storage resources are utilized to store raw telemetry data and ML inference results. These resources also allow for retraining of models and data analysis closer to the data source, reducing latency and bandwidth usage.

End-to-End Workflow Integration

- **Data Lifecycle Management:** The workflow seamlessly integrates data ingestion, processing, storage, and analysis across edge and cloud environments. Anomalies are detected and addressed both in real-time at the edge and through deeper analysis in the cloud.
- **Feedback Loop:** The system's design creates a feedback loop where insights gained from cloud analysis influence the operations at the edge, ensuring continuous improvement and adaptation of the IoT system.