Monitoring Streaming Data with Machine Learning

Identify and act on deviations from forecasts in near-real-time

This architecture enables customers to monitor streaming data and compare it in near-real-time to a machine-learned forecast, raising an incident or alarm if actual performance deviates significantly from the forecast.

---

**1. Data Collection**

- Amazon Kinesis Data Streams
- Amazon Kinesis Data Firehose
- Lambda function

**2. Data Preparation, Model Training, and Forecast Generation**

- Amazon Athena
- S3 Bucket
- Amazon SageMaker

**3. Near Real Time Aggregation and Validation**

- Amazon Kinesis Data Analytics
- Lambda function

**4. Monitoring, Alerting, and Remediation**

- Amazon CloudWatch
- AWS Systems Manager Incident Manager

---

1. Data is collected from multiple data sources across the enterprise and the edge using Amazon Kinesis Data Streams' many SDKs with support for languages like Java, .NET, C++, Python, Javascript, and others.

2. Data persists and is sent to Amazon Simple Storage Service (Amazon S3) by Amazon Kinesis Data Firehouse. AWS Lambda can be used to enrich data prior to storage in Amazon S3.

3. Initial data preparation and aggregation is performed using Amazon Athena. Prepared and aggregated data is stored in Amazon S3.

4. Amazon SageMaker is used to train a forecasting model and create predictions of future behavior. These can be predictions for either statistical descriptions (for example sample counts and standard deviations) or business-oriented aggregations (for example transaction values). The predictions are stored in Amazon S3.

As new data arrives, it is aggregated and prepared in near-real-time by Amazon Kinesis Data Analytics. The resulting prepared data is compared to the previously generated forecast.

5. Using another AWS Lambda function, the forecast and actual values are written as metrics to Amazon CloudWatch.

6. When actual values deviate significantly from the forecast, a CloudWatch alarm triggers an incident in AWS Systems Manager Incident Manager to trigger an investigation or remediation.