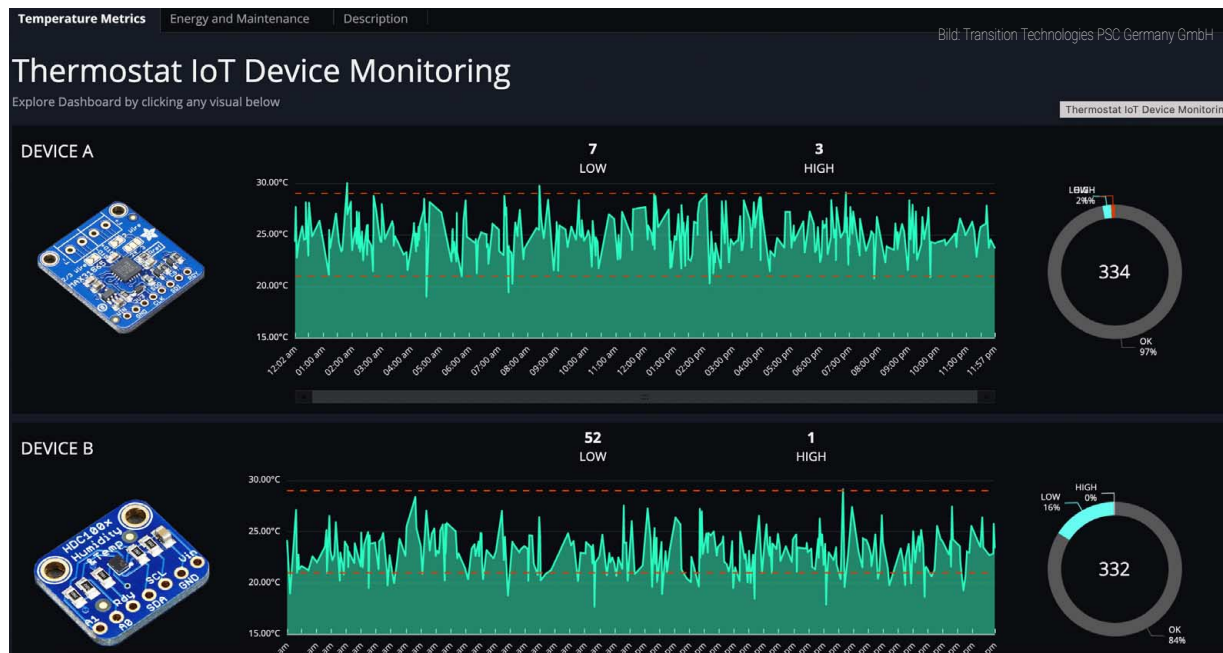


IIoT-Setup am Beispiel der AWS-Plattform

Wie Maschinendaten auf das Dashboard kommen



Ein mit AWS-Bordmitteln erstelltes Dashboard

Über den Nutzen des Industrial Internet Of Things (IIoT) wird viel geschrieben, über die konkrete Einrichtung kaum. Dieser Artikel veranschaulicht den ersten Aufbau einer kosteneffizienten Serverless-Pipeline für die Langzeitdatenspeicherung und Datenanalyse auf der AWS-Plattform.

Die Anwendungsmöglichkeiten, die unter dem Schlagwort Industrie 4.0 zusammengefasst werden reichen vom vermeintlich einfachen Monitoring der wichtigsten Betriebsdaten, über die KI-gestützte Analyse von kriti-

schen Kennzahlen bis zu automatisierten Steuerungssignalen für zentrale Produktionsprozesse. Doch wo sollen Unternehmen einsteigen? Dieser Beitrag stellt erste Lösungsbausteine bereit. Die Herausforderungen lassen sich dabei in drei folgenden Kategorien einteilen:

- Konnektivität
- Datenanalyse
- Visualisierung der Ergebnisse

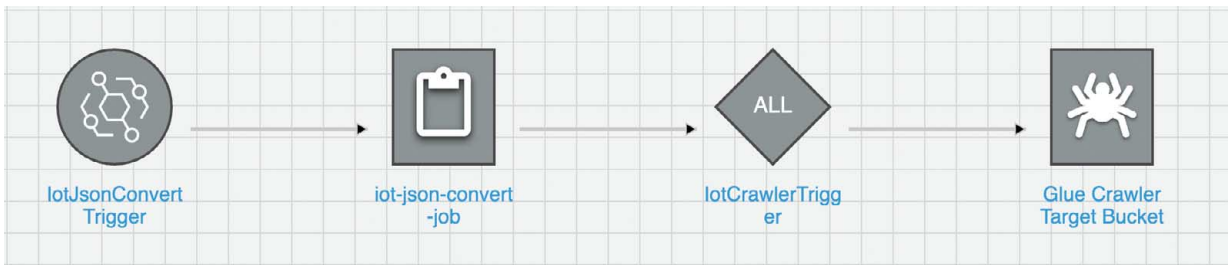
Konnektivität zwischen lokalen Maschinen und der Cloud

Wie konnektieren Unternehmen ihre lokalen Maschinen mit der Cloud? Am Beispiel von Amazon Web Services (AWS) zeigen verschiedene Möglichkeiten:

- Ein bestehender MQTT-Broker leitet die Nachrichten direkt an AWS IoTCore weiter, dazu kann dieser beispielsweise als IoT Thing angelegt werden und sich per Zertifikat authentifizieren.
- Devices können auch direkt mit IoTCore verbunden werden.
- Greengrass als OT-Software auf dem Shopfloor ist die Standardsoftware von AWS, um eine Verbindung zwischen Shopfloor-Devices und der AWS-Cloud herzustellen. Sie wird als Container bereitgestellt und über die AWS-Konsole administriert.

Zu beachten ist, dass ein einzelner Greengrass-Container nicht darauf ausgelegt ist, alle Verbindungen für eine gesamte Firma zu verwalten, sondern dass mehrere Container abhängig von der Anzahl

Bild: Transition Technologies PSC Germany GmbH



Visualisierung des Glue-Workflows zum Aufbereiten der Rohdaten und Indizieren der Ergebnisdaten nach der Transformation.

zu erfassenden Datenpunkte benötigt werden. Im Folgenden wird von einer Anbindung über IoTCore ausgegangen.

Zwischen kalten und heißen Daten

Die Telegramme werden in der Cloud empfangen, sodass die tatsächlichen Pain Points, wie der Energieverbrauch oder die Früherkennung fehlerhafter Bauteile, adressiert werden. Im Wesentlichen muss zu Beginn zwischen zwei unterschiedlichen Arten der Analyse unterscheiden:

- Echtzeit-Prozessierung von Daten (hot-data),
- Langzeit-Betrachtung von Daten (cold-data).

Aufgrund der höheren Komplexität und meist höheren Kosten, wird mit dem cold-data-Pfad für die Langzeitanalyse begon-

nen. Das weitere Vorgehen hängt von den individuellen Gegebenheiten und Anforderungen ab, da sich die Anwendungen durchaus in Durchsatz, Latenz und Komplexität unterscheiden können. Zu Beginn eines Projektes gilt es, die Komplexität gering zu halten, da bei der Einführung einer IoT-Plattform selten die Grenzen der 'Standard'-Services erreicht werden. Natürlich sollte die Anwendung mit wachsender Datenmenge skalieren.

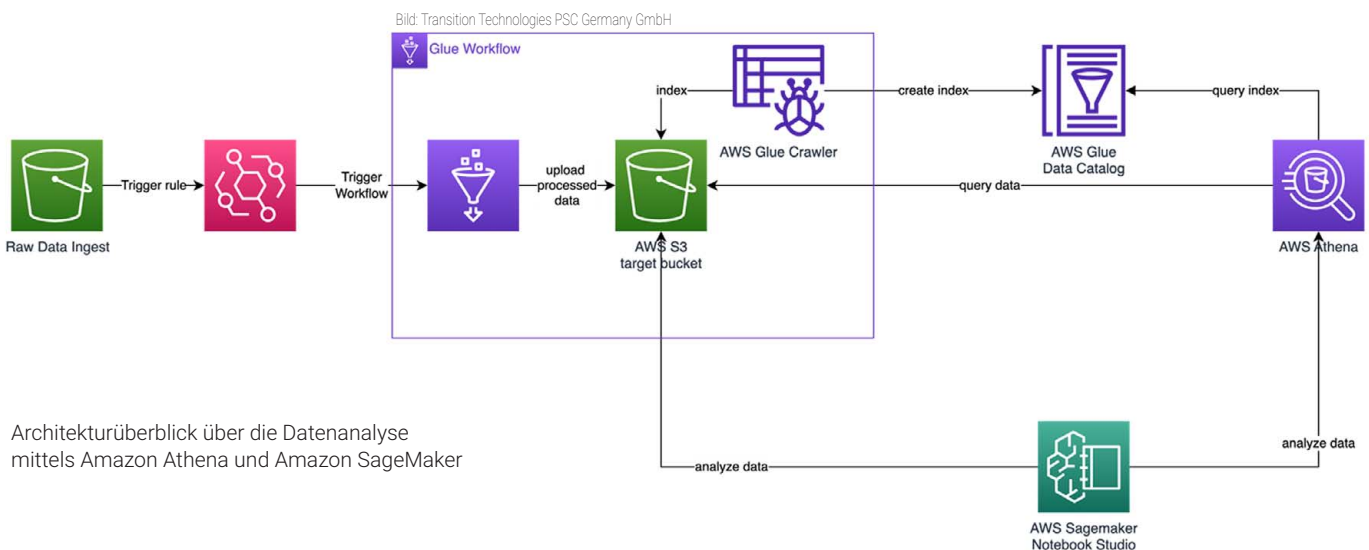
Regeln für die kalten Daten

Wenn die Daten per IoT-Core entgegen genommen werden, kann das weitere Aussteuern der Daten über IoT-Rules umgesetzt werden. Diese ermöglichen es anhand von Topics und einigen Body-Informationen, unterschiedliche Folgeaktionen auszulösen. Beispielsweise kann anhand der Topic-Struktur

definiert werden, welche Daten in den hot- und welche in den cold-data-Pfad fließen. Dieser Artikel beschränkt sich auf den Aufbau des cold-data-Pfades. Dazu wird in der IoT-Rule eine Aktion definiert, die ankommenden Daten direkt in Kinesis Data Firehose übergibt. Dieser Service puffert und aggregiert die Daten, bis eine gewisse Menge (etwa ein Megabyte) erreicht ist oder nach einem Zeitintervall, um diese dann an den Objektspeicher (S3) weiterzugeben. So wird eine stabile und skalierbare Datenstrecke aufgebaut, über welche die Rohdaten etwa im Json-Format entgegengenommen und abgespeichert werden können.

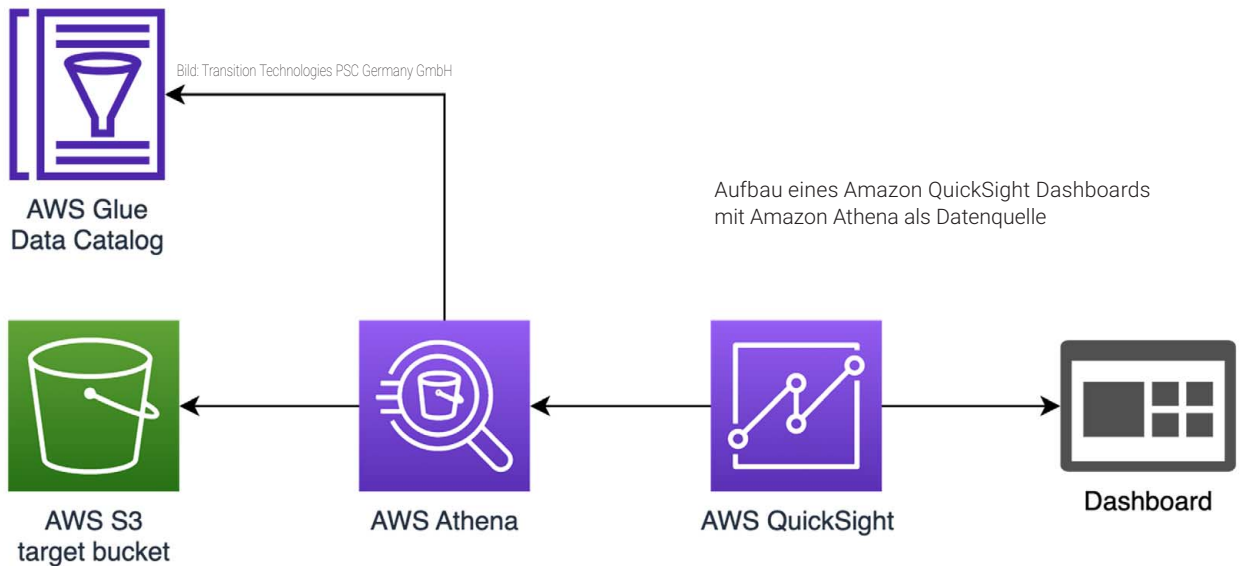
Datenaufbereitung und -analyse beginnt

Für die weitere Aufbereitung und Weiterverarbeitung der abgelegten Rohda-



Architekturüberblick über die Datenanalyse mittels Amazon Athena und Amazon SageMaker

Maschinenanbindung



ten bietet die AWS-Cloud unterschiedliche Möglichkeiten. Ein Ansatz ist eine ETL (Extract-Transform-Load)-Pipeline im AWS Serverless-Dienst Glue. Über diesen Dienst kann eine Pipeline aufgebaut werden, um die Datenpunkte über ein Python-Skript zu bereinigen, erste Berechnungen durchzuführen und die Ablage in einem platzsparenden und effizient abfragbaren Format durchzuführen. Als Standard hat sich das spaltenbasierte Parquet-Format durchgesetzt. Nach dem eigentlichen Job können die Ergebnisse durch einen Crawler für die effiziente Abfrage indiziert werden. Dadurch ergeben sich zahlreiche Integrations- und Analysemöglichkeiten, wie etwa durch den Dienst Amazon Athena. Dieser ermöglicht eine SQL-basierte Abfrage der Daten und kann dadurch als Grundstein für weitere KI-gestützte Analysen genutzt werden. Auch die explorative ad-hoc-Analyse der Daten ist über

ein Jupyter-Notebook möglich. Mit diesem Aufbau können Unternehmen bereits Nutzen erzielen, indem sie beispielsweise die Energieverbräuche in der Produktion erfassen und automatisiert Verbesserungspotentiale ermitteln.

Visualisierung der Ergebnisse

Der offene Ansatz der Architektur vereinfacht die Integration in bereits bestehende Systemlandschaften. So können etwa wiederkehrende Abfragen über ein API-Gateway bereitgestellt und über http abgefragt werden. Auch eine automatische Synchronisation der Daten auf mobile Endgeräte für Servicetechniker, Instandhalter oder Außendienstler ist durch eine standardisierte Schnittstelle wie AWS AppSync umsetzbar. Auch der Datentransfer per Schnittstelle in bestehende Business Intelligence-Anwendun-

gen wie Power BI ist möglich. Alternativ bietet AWS selbst mit seiner BI-Lösung QuickSight ein leistungsfähiges Werkzeug, um die Daten aus der Fabrik zu analysieren und die Ergebnisse in Dashboards zu visualisieren.

Bereit für die Arbeit mit Echtzeit-Daten

Sind alle Arbeiten abgeschlossen, verfügen Unternehmen über den ersten Aufbau einer kosteneffizienten Serverless-Pipeline für die Langzeitdatenspeicherung und Datenanalyse auf der AWS-Plattform. Dieser bildet die Basis für weitere Projekte. Nach einem erfolgreichen Proof of Concept lässt sich im nächsten Schritt die Verarbeitung von Live-Daten angehen. Diesem Thema ist demnächst ein eigener Artikel gewidmet. ■

[ttpsc.com](https://www.ttpsc.com)



Autor

Torsten Hopf ist Senior Business Analyst bei Transition Technologies PSC