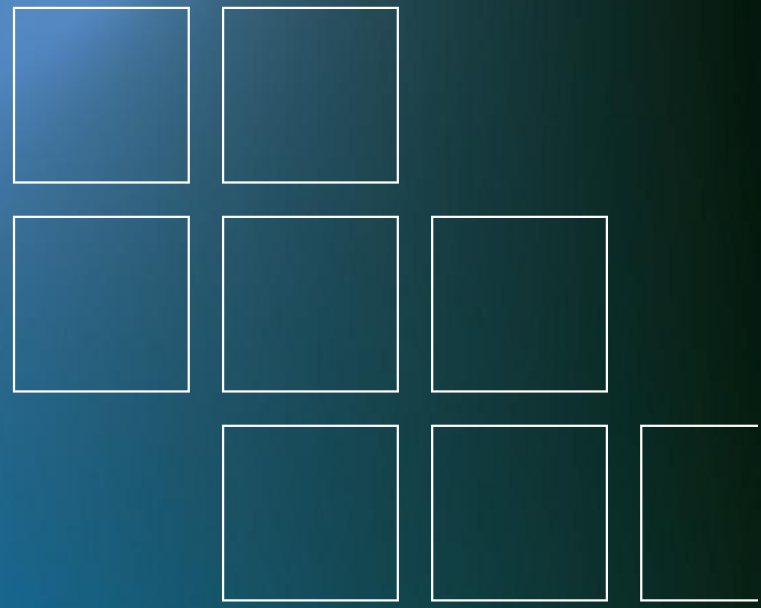
controlware

Smart Data Management: Shaping und Optimierung mit Splunk Edge Processor und S3

Splunk Public Sector Summit



Ihre Ansprechpartner

Nahbar, Direkt & Erreichbar

Jorg Veit

Senior Technical Consultant Analytics Engineering

E-Mail jorg.veit@controlware.de

Daniel Seifert

Teamlead Analytics Engineering

E-Mail daniel.seifert@controlware.de

Ausgangssituation & Zielbild

Ausgangssituation

Es kommen im mehr Daten an

Neue Datenquelle mit erhöhtem Volumen

Bestehende Datenquellen verbrauchen höhere Volumina

Regeln für Compliance werden immer strenger

Aufbewahrungszeiten müssen erhöht werden

Log Level / Detailierung muss angepasst werden

Begrenztes Budget bzw. Splunk Lizenzvolumen

Budget wird nur gering oder gar nicht erhöht

Lizenzvolumen stagnieren

Ausgangssituation & Zielbild

Zielbild

Alle relevanten Daten und Felder in Splunk indizieren

- Gesteigerten Anforderungen gerecht werden
- Neu Datenquellen anbinden

Gleichzeitig vollständiges Rohdatenarchiv für Compliance & Forensik

- Aufbewahrung aller ankommenden Daten in Ihrem Ursprungs-Format
- Schnelle Wiederherstellungsmöglichkeit der Rohdaten

Volle Kontrolle

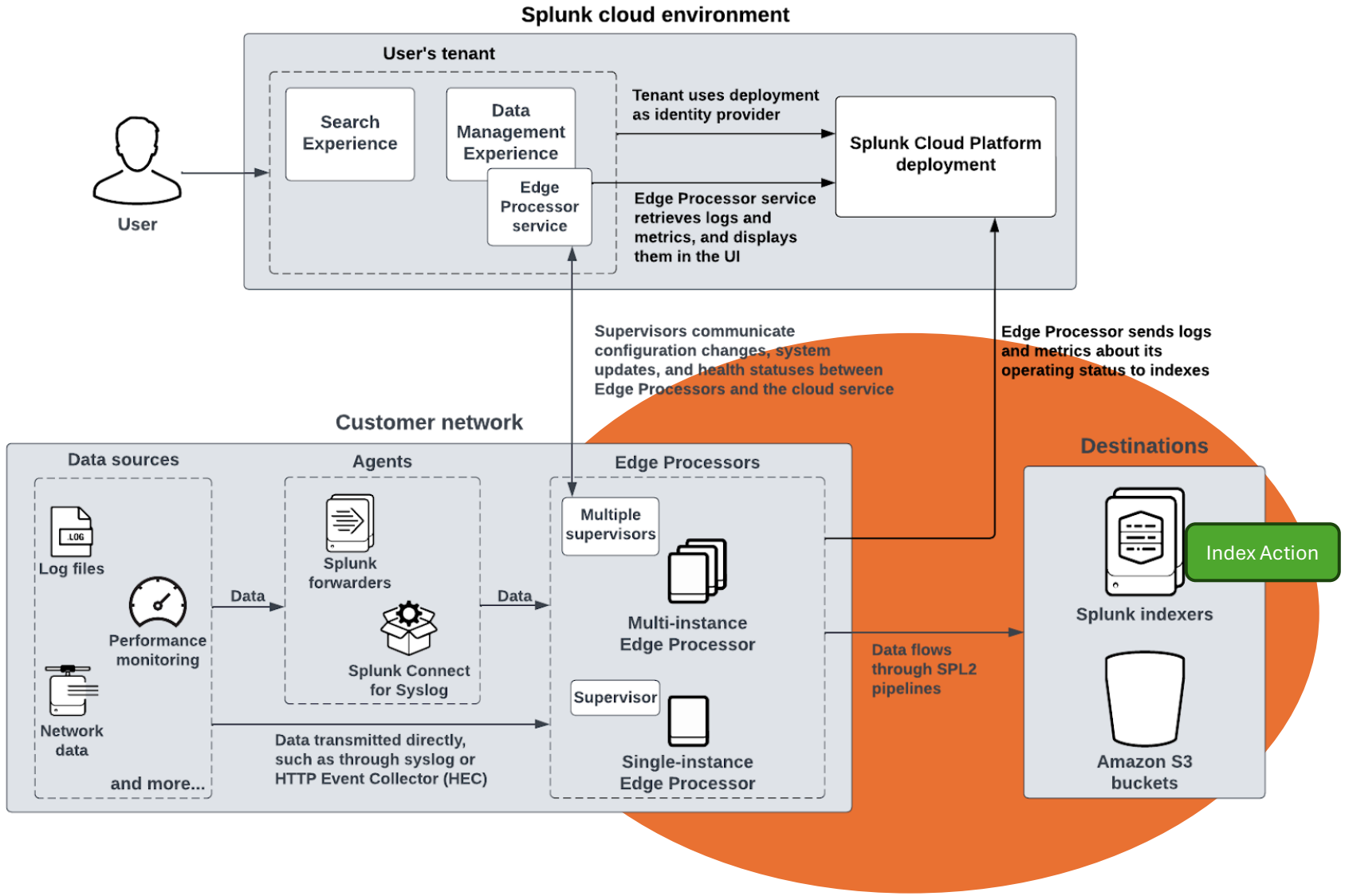
- Keine Beschaffung neuer Lizenzen
- Keine Einführung neuer Software
- Kein Ausleiten an Cloud Services (onPrem S3)

Ausgangssituation & Zielbild

Zielbild

Splunk Edge Processor	Splunk Enterprise	Splunk Custom Command	Splunk Index Actions
• Annahme der Daten • Datenvorverarbeitung und shaping	• Indizierung relevanter Daten	• Durchsuchen der S3 Daten • Wiederherstellung S3 Daten	• Weiterleitung an S3 kompatiblen Speicher

Herausforderungen:

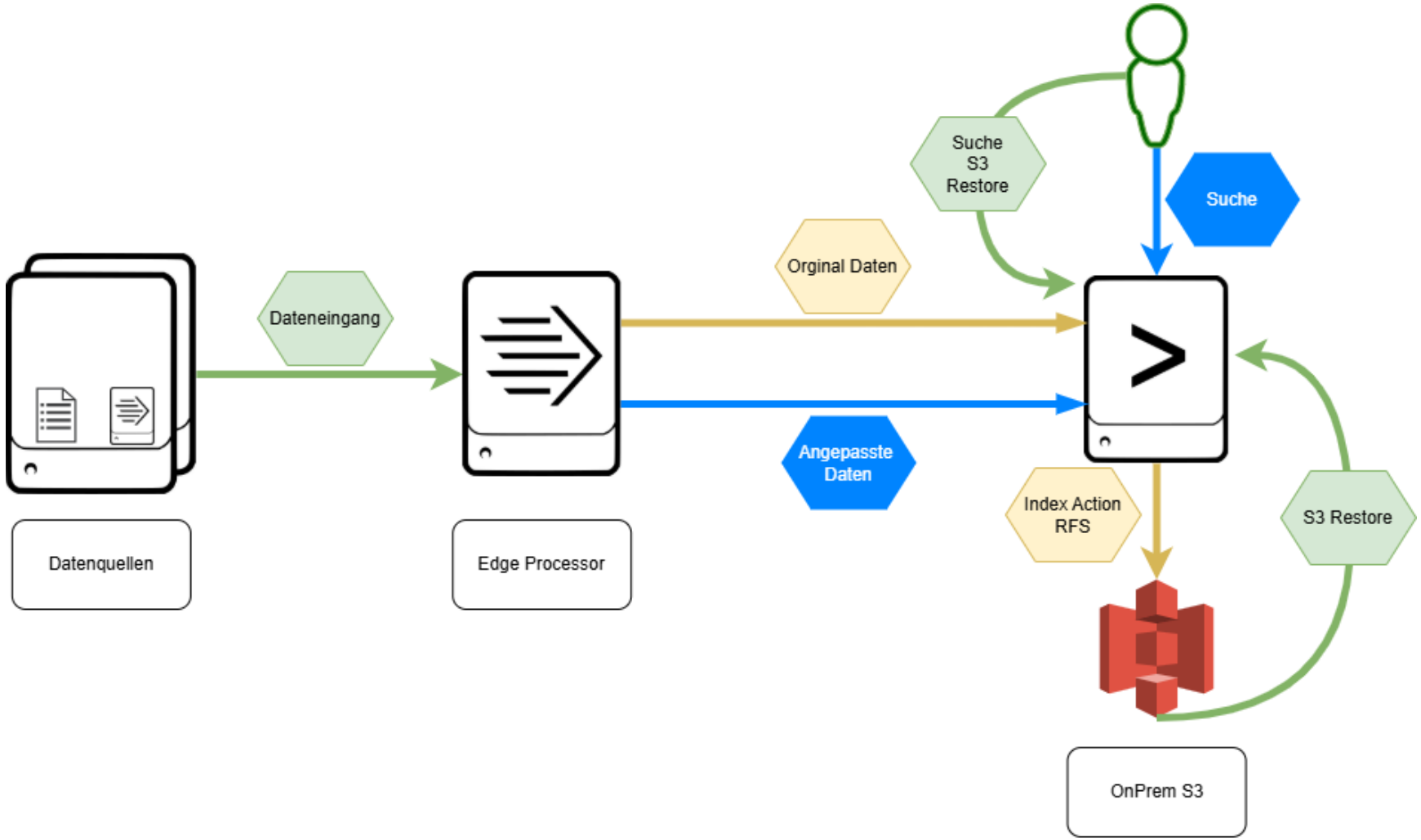


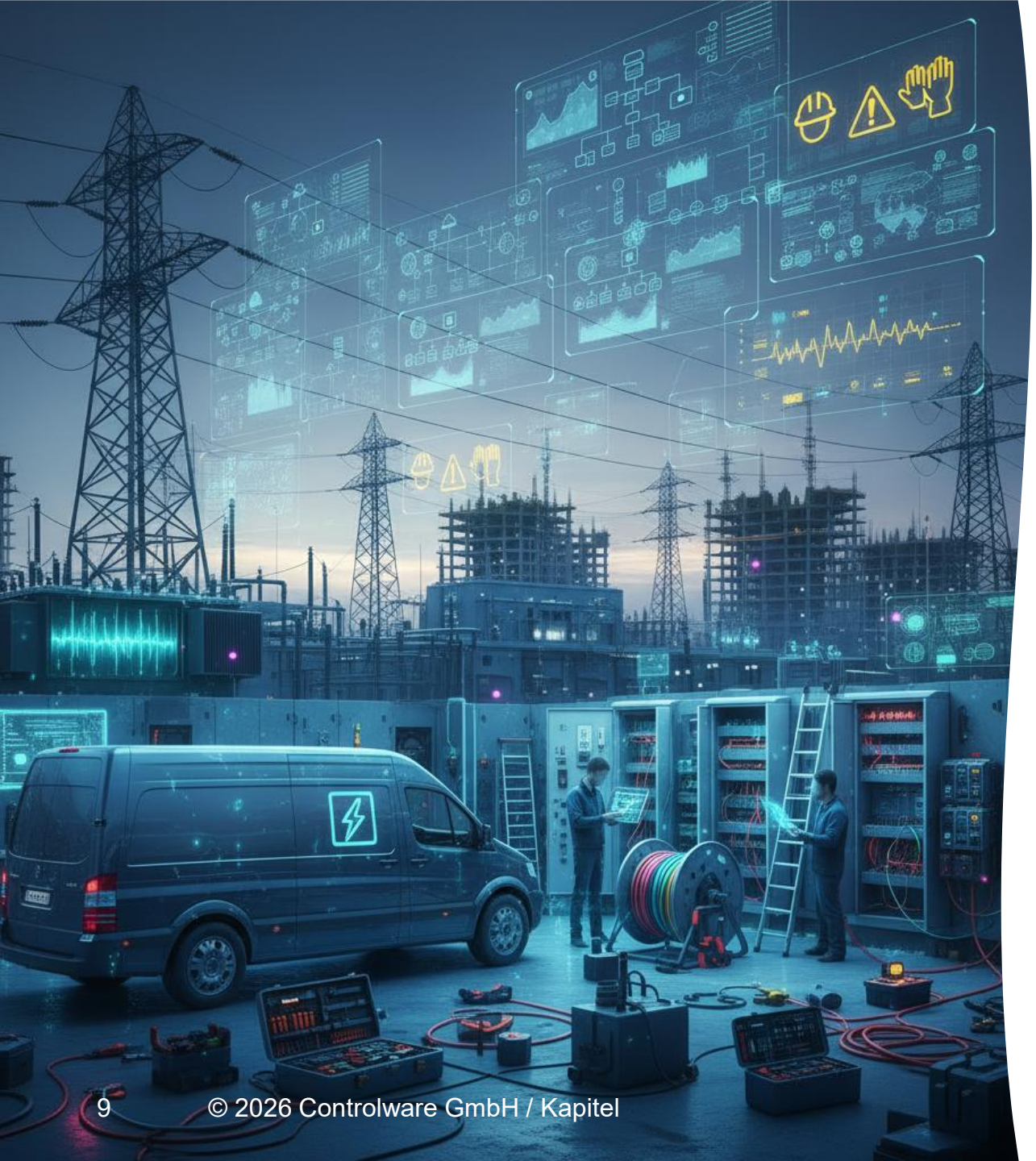
controlware

Herausforderungen

- Edge Processor kann aktuell keine Daten an OnPrem S3
- Ablagestruktur der S3 Daten durch Splunk (Pfad)
- Index Actions finden nach dem Edge Processor statt
- Wiederherstellung sollte einfach für den Anwender sein
- Welche Daten brauchen wir wirklich?
  - Filtern von technischen „Noise“-Events (z. B. Health-Checks)
- Welche Felder sollen bleiben?
  - Konzentration auf Felder, die für Use Cases relevant sind
- Welche Informationen sind sensibel?
  - Maskierung von Benutzernamen, IP-Adressen, E-Mail-Adressen etc.

Lösungsmöglichkeit:





controlware

Lösungsmöglichkeit

- Ein Event kommt beim Edge Processor an
- Es wird einmal geparst und verstanden
- Dann passieren zwei Dinge parallel:
  - **Kopie 1 (für Splunk):**
    - Entrümpelt und aufbereitet
    - Nur die wesentlichen Felder bleiben
    - Event wird in Splunk indiziert
  - **Kopie 2 (für S3):**
    - Unverändert als Rohdaten
    - Wird im S3-Bucket abgelegt
- User kann auf zwei arten auf die Daten zugreifen
  - Splunk: Suche in den angepassten Daten
  - Splunk: Suche über ein Custom-Command
    - Direkt in S3 oder
    - Restore der Daten anfordern

controlware

Details
Edge Processor

- Universal Forwarder sendet Roh-Events per S2S an den Splunk Edge Processor
- Im Edge Processor wird der Datenstrom dupliziert (zwei parallele Pipelines)
  - **Stream 1 (transformiert):**
    - Events werden beschnitten
    - Weiterleitung der optimierten Events an Splunk
  - **Stream 2 (roh):**
    - Events bleiben unverändert
    - 1:1-Weiterleitung per S2S an Splunk

controlware

Details
Ingest Action

- Der Splunk Edge Processor sendet vorverarbeitete Daten an eine Index Action auf einem dedizierten Port.
- Die Index Action übernimmt die Daten, führt keinen Indexierungsvorgang in Splunk durch.
- Stattdessen werden die Events über das Remote File System (RFS) als Dateien geschrieben.
- RFS ist mit einem on-prem S3-kompatiblen Storage verbunden; die Daten landen direkt im S3 Bucket.

controlware

Details
Custom Command

- Custom Command liest S3-Objekte ein und schreibt deren Metadaten in den Splunk KV Store (Index, Host, Source, Sourcetype)
- Zuerst wird nur der KV Store durchsucht: schnelle Identifikation relevanter S3-Objekte anhand von Zeit, Source, etc.
- Für die im KV Store identifizierten S3-Objekte lädt das Custom Command bei Bedarf die Rohdaten aus dem S3-Bucket
- Bereitstellung in Splunk (temporär oder in definiertem Re-Index) zur Analyse mit gewohnter SPL

Vielen Dank für Ihre Aufmerksamkeit.
Thank you very much for your attention.

controlware