
Success Story

Audi Electronics
Venture GmbH



→ Audi Electronics Venture GmbH

Die Fakten auf einen Blick:

Aufgaben

- ⊕ Expertenunterstützung beim Aufbau eines DC/OS-Clusters auf Amazon Web Services (AWS)*
- ⊕ Aufbau der Fast-Data-Plattform SMACK mit DC/OS
- ⊕ Entwicklung von Ingest- und Digest-Applikationen mit Scala

Vorgehen

- ⊕ Wissenstransfer der Spezialkenntnisse der codecentric AG in den Bereichen AWS, Mesosphere DC/OS, SMACK und Scala in die Audi Electronics Venture GmbH (AEV)*
- ⊕ Gemeinsame Hands-on-Implementierung im Rahmen eines Scrum-Projekts

Projektziele

- ⊕ Aufbau von zwei DC/OS-Clustern auf AWS mit insgesamt 50 Instanzen für Zero-Downtime-Deployments
- ⊕ Installation der Fast-Data-Plattform SMACK in einen DC/OS Cluster zur Realtime-Verarbeitung von Sensordaten
- ⊕ Schulung der AEV in den Bereichen SMACK, Scrum und Scala

Ergebnisse

- ⊕ Zwei Analyse-Cluster (mit jeweils 25 Maschinen), die mit Hilfe von Jenkins-Jobs in circa drei Minuten in AWS erstellt werden
- ⊕ Der Betriebszustand ist nach weiteren 30 Minuten erreicht und die Fast-Data-Plattform kann für Analysen vollständig genutzt werden
- ⊕ Aktuell startet die Plattform ca. 750.000 Spark-Tasks am Tag
- ⊕ Aufbau des Elastic Stack für die verteilte Logdaten-Verarbeitung (20GB Logs/Tag)



Innovative Elektronik bildet das Fundament für das intelligente Auto von morgen. Die Audi Electronics Venture GmbH (AEV)* gestaltet die Zukunft der Mobilität aktiv mit. Als 100-prozentige Tochtergesellschaft ist sie Impulsgeber und Vorentwicklungspartner für die AUDI AG*. Dank eines großen Entwicklungsnetzwerks bringt die AEV Experten zusammen, die Innovationen und Technologien bis zur Marktreife entwickeln. Damit schafft und sichert sie den Zugang zu Schlüsseltechnologien für Audi.

Die AEV war auf der Suche nach einem Partner, der sie dabei unterstützt die Fast-Data-Plattform SMACK zur performanten Datenanalyse einzuführen, um Fahrzeugdaten mit Hilfe von Streams zu verarbeiten. Der SMACK Stack ist ein Ensemble aus verschiedenen Tools, die miteinander zu einer Big-Data-Plattform integriert werden. Der Name SMACK ist ein Akronym aus den Anfangsbuchstaben der Tools Apache **S**park, Apache **M**esos, **A**kka, Apache **C**assandra und Apache **K**afka.

Betrachten wir im Folgenden die einzelnen Komponenten:

- Apache Spark unterstützt sowohl Batch-als auch Stream-Verarbeitung und dient dazu eine Lambda-Architektur umzusetzen.
- Apache Mesos ist ein Scheduling-Framework für Cluster und Rechenzentren. Anwendungen, wie Spark, Akka, Cassandra und Kafka, werden durch Mesos in das Cluster vollautomatisch ausgeliefert.
- Akka ist eine Implementierung des Aktoren-Modells und erlaubt das Schreiben von reaktiven Anwendungen, die hoch-verteilt und -performant sind.
- Apache Cassandra ist eine spaltenorientierte Datenbank. Sie kann über verschiedene Rechenzentren verteilt werden. Die Datenbank skaliert nahezu linear zur Anzahl der Maschinen im Rechnernetz. Die Datenbank hat eine sehr gute Performance, auch bei gemischten Schreib- und Lesevorgängen.
- Apache Kafka ist ein verteilter, skalierbarer und verlässlicher Message Broker für große Datenmengen. Kafka partitioniert die Daten und speichert sie in einem Append-Only-Log, so kann Kafka ohne Weiteres Terabytes an Daten vorhalten – ohne Performanceeinbußen.

Um die unterschiedlichen SMACK-Komponenten auf die unterschiedlichen Rechner in einem Cluster zu installieren, eignet sich DC/OS hervorragend. DC/OS ist ein verteiltes Betriebssystem für



Rechenzentren und basiert auf dem Apache-Projekt Mesos. DC/OS verfügt über ein Repository, aus dem sehr einfach Komponenten wie Spark, Cassandra und Kafka installiert werden können.

Gemeinsam mit der AEV wurde innerhalb von nur wenigen Wochen ein vollständig lauffähiges DC/OS-Cluster auf AWS bereitgestellt. Dabei wurde auf eine komplette Automatisierung der Plattform mit Terraform gesetzt. Terraform ist ein Werkzeug, mit dem sich die Infrastruktur bzw. ein ganzes Rechenzentrum in Code beschreiben und verwalten lässt. Der Terraform Code kann ausgeführt werden, um in AWS ein komplettes Rechenzentrum inklusive aller Komponenten, wie Netzwerke, Instanzen, Routing-Tabellen und Loadbalancer zu erstellen.

Die Cluster-Erstellung kann mit einem Knopfdruck in einem Jenkins gestartet werden. Außerdem erfolgt im Jenkins eine Auditierung, wann, wer, welchen Job gestartet hat. Nach dem Starten des Jenkins-Job dauert es circa 30 Minuten bis der komplette SMACK Stack inklusive aller Ingest- und Digest-Applikationen installiert ist.

Logausgaben sind immer noch eine wichtige Quelle für Informationen über Anwendungen. Als Überwachungstool wurde deshalb der Elastic Stack eingeführt, um die Logdaten des verteilten Systems analysieren zu können. In einem SMACK-Cluster gibt es verschiedene Quellen wie Spark, Mesos, Akka, Cassandra und Kafka, an denen Logausgaben entstehen können.

Der Elastic Stack besteht aus Tools wie Elasticsearch, Logstash und Kibana. Elasticsearch wird für die Ablage der Logdaten benutzt. Außerdem ist es für die Volltextsuche sehr gut geeignet. Logstash kann Logdaten parsen und aus verschiedenen Quellen wie Dateien, aber auch Messaging-Systemen oder TCP/IP-Ports lesen und schreibt diese nach Elasticsearch. Mit Kibana können die Daten durchsucht oder auch grafische Auswertungen erzeugt werden. Das Zusammenspiel all dieser Tools macht es letztlich einfach Logdateien von mehr als 50 Rechnern einfach an einer zentralen Stelle zu analysieren.

Es wurden von der codecentric AG alle nötigen Jenkins-Jobs mit der Jenkins DSL geliefert, um das Deployment des DC/OS SMACK-Clusters vorzunehmen. Die AEV-Mitarbeiter wurden entsprechend geschult, um Letztere auch nach dem Experteneinsatz der codecentric AG nicht mit dem Betrieb des Clusters allein zu lassen.



Außerdem wurde zusammen mit der AEV exemplarisch ein Realtime Streaming Usecase erstellt, der Sensordaten verarbeitet und aufbereitet. Die Visualisierung erfolgt in einer Javascript-Anwendung. Um eine weitere Visualisierung zu bekommen, wurde weiterhin Zeppelin eingeführt. Zeppelin ist eine in den Browser integrierte Benutzeroberfläche, die über Notebook-ähnliche Fähigkeiten für Analysten und Datenwissenschaftler verfügt, damit sie ihre Daten interaktiv untersuchen und anspruchsvolle Datenanalysen durchführen können. Zeppelin kommt auch bei Amazon Web Services, JuJu, Hortonworks und Twitter zum Einsatz, um Daten zu analysieren.



Stefan Siprell, Head of Data
stefan.siprell@codecentric.de

„Dank der hervorragenden Expertise der codecentric im Bereich Automatisierung und Big-Data-Technologien ist es gelungen, binnen kürzester Zeit einen Fast Data Stack in Betrieb zu nehmen. Ein neues Release der entstandenen Plattform kann dank der starken Automatisierung binnen kürzester Zeit zur Verfügung gestellt werden.“

David Schmitz, Lead Big Data Architekt der Audi Electronics Venture GmbH

* Markennamen und geschützte Warenzeichen sind Eigentum ihrer jeweiligen Inhaber. Die Nennung von Markennamen und geschützter Warenzeichen hat lediglich beschreibenden Charakter. Genannte Marken stehen in keinerlei Partnerschaft oder Kooperation zur codecentric AG. Die Angabe der Marken erfolgt durch den jeweiligen Autor/Nutzer. Irrtümer vorbehalten.