Logmanagement: aktuelle Treiber, Lösungen und Praxis

Datenzentrale für DevOps

Es ist noch gar nicht lange her, da interessierte sich in den meisten Unternehmen allerhöchstens der IT-Betrieb für ein systematisches Logmanagement. Im Zuge der aktuellen technologischen Entwicklungen wie Cloud Computing und Internet der Dinge und den damit verbundenen regulativen Anforderungen, neuen Projektabwicklungsmethoden und IT-Organisationsformen wird es für Unternehmen immer wichtiger, den Zustand eines Systems genau zu kennen und schnelle Fehlerdiagnosen zu erstellen. Deshalb erwarten viele CIOs heute ein übergreifendes Logmanagement. Welche Funktionen erfüllt das Logmanagement in den Unternehmen? Welche konkreten Treiber gibt es? Die Autoren stellen Lösungen für ein effektives Logmanagement vor und zeigen, wie es sich in der Praxis umsetzen lässt.

AUTOREN: RICHARD ATTERMEYER UND WALDEMAR KAUS

Immer mehr Anwendungen werden in privaten oder öffentlichen Clouds betrieben [1]. Die Anwendungen laufen also nicht mehr auf realen Servern, sondern in virtuellen Maschinen oder Containern. Mit diesem Trend zu leichtgewichtigeren Ablaufumgebungen verändert sich auch die Lebensdauer der Umgebungen. Umgebungen werden im Extremfall direkt neu erstellt, sobald sich die Konfiguration ändert, und die alte Version wird gelöscht.

Ohne ein zentrales Logmanagement gehen die Logs der Anwendungen also verloren und mit ihnen auch eine wichtige Informationsquelle: Dabei werden die Logs gebraucht, um Development und Operations bei der Untersuchung von Fehlern zu unterstützen.

Doch die Hilfe bei der Fehlersuche ist nicht der einzige Nutzen einer zentralen Logsammlung. Je mehr Firmen sich in Richtung Cloud orientieren, umso kleinteiliger werden ihre Applikationen betrieben, bis am Ende jede Miniapplikation in einem eigenen Container läuft. Doch gerade in verteilten Umgebungen ist es für den Betrieb und die Softwarewartung schwierig, Anfragen über verschiedene Knoten zu verfolgen und in den einzelnen Applikationslogs nach Hinweisen zu suchen. Traditionell müssen sich die Administratoren dazu an jedem Knoten anmelden und die Logs durchsuchen oder sie manuell zusammentragen. Das ist zeitaufwändig und skaliert nicht.

LOGS ALS DATENQUELLE FÜR SICHERHEIT UND COMPLIANCE

Mit dem Vormarsch des Internets der Dinge oder den Cyber Physical Systems kommt dem zentralen Logmanagement und seiner Analyse eine neue Bedeutung zu. Während die Anzahl der Geräte, die Loginformationen schreiben, nun stetig zunimmt, werden auch Logs als Datenquelle immer interessanter, da sie nach bestimmten Mustern untersucht werden können. Insbesondere wenn die Logdaten von einer zentralen Event Processing Engine verarbeitet werden, ergeben sich neue Chancen, zum Beispiel für die proaktive Wartung von Industrieanlagen.

Aber auch bei der Korrelation von Firewall- und Application-Logs können sich Indizien ergeben, die bele-

gen, ob ein Angriff auf eine Applikation stattfindet oder stattgefunden hat. Neben der Aufklärung von Angriffen kann eine automatische Auswertung dabei helfen, Angriffe früher zu erkennen und zu verhindern.

AGILE ENTWICKLUNG UND DEVOPS

Bei der Art und Weise, wie Software entwickelt wird, ist seit Jahren ein Trend in Richtung agiler Methoden zu beobachten. Damit agile Methoden ihren Mehrwert generieren können, verkürzen sie unter anderem die Feedbackschleife. Deshalb werden immer mehr Test- und Build-Umgebungen benötigt, auf denen automatische oder manuelle Tests ablaufen und auf denen Qualitätssicherung betrieben werden kann.

Früher wäre das zu kostspielig gewesen, deshalb war auch nur eine begrenzte Anzahl an Testumgebungen üblich. Mit dem Trend hin zu virtuellen Umgebungen und Containern sind diese Hürden überwunden. Im Rahmen von Continuous-Delivery-Ansätzen [2] bekommt heute jeder Integrations-Build einer Software eine eigene Ablaufumgebung. Auf diese Weise kann die Qualitätssicherung auf Knopfdruck eine Umgebung aufbauen und auch wieder abreißen. Ein typischer Ablauf sieht vor, dass ein Tester eine entsprechende Umgebung für seine Tests startet, seine Tests durchführt, Bugs meldet und dann die Umgebung wieder löscht.

Wie aber sollen Entwickler Bugs nachvollziehen, wenn die Umgebung nicht mehr existiert? Dazu müssten sie außerhalb der Ablaufumgebung auf die Logdaten zugreifen können.

In vielen Unternehmen gibt es bereits eine engere Kooperation zwischen Development und Operations, bekannt unter dem Stichwort DevOps. Die engere Zusammenarbeit soll unter anderem dabei helfen, schneller auf Problemberichte von Kunden reagieren zu können. Dazu benötigen Entwickler wichtige Daten. Diese bekommen sie heute aber oft noch nicht auf dem direkten Weg. Die meisten Administratoren stellen die notwendigen Logdateien selbst zusammen, um zu vermeiden, dass Entwickler auf die Produktion zugreifen. Das ist aufwändig. Besser wäre es, Entwicklern den direkten Zugriff auf die entsprechenden Logdaten zu gewähren.

Moderne Logmanagementlösungen ermöglichen einen direkten Zugriff. Mit ihnen können Entwickler mit einem Client oder mit einem Browser ungehindert auf die Logdaten zugreifen, diese durchsuchen und korrelieren.

EIN BLICK INS LOG HILFT WEITER

Bei vielen aktuellen Themen kann es hilfreich sein, regelmäßig ins Log zu schauen: Sachbearbeiter melden sporadische Fehler in der zentralen Unternehmensanwendung. Die automatisierte Testsuite läuft aus unbekannten Gründen nicht durch. Aus dem Second-Level-Support werden weitere Informationen zum gemeldeten Fehlerbild angefragt. Das Monitoring meldet ein mit Logdateien überquellendes Dateisystem. Die Entwickler würden gern das Verhalten der Anwendung in Produktion und Testsystem vergleichen. Die Liste ließe sich endlos fortführen ... Ein einfacher Blick ins Log würde hier helfen.

Doch stattdessen beginnt an dieser Stelle für den zuständigen Analysten eine kleine Odyssee: Über welche Systeme sind die Logs verteilt? Auf welchem Clusterknoten wurden die gesuchten Informationen abgelegt? Wie lauten die Dateipfade? Habe ich Zugriff darauf? Wie vergleiche ich geschickt unterschiedliche Zeiträume? Wie korreliere ich die Logs aus unterschiedlichen Anwendungen in einem verteilten System? Welche Logausgaben führen dazu, dass die Logfiles ins Unermessliche anwachsen? Wie fasse ich rollierte und archivierte Logdateien zusammen? Selbst ein geübter Kommandozeilenbenutzer stößt hier mitunter an seine Grenzen.

Diese und ähnliche Komplikationen führen dazu, dass Logdateien bei Fehleranalysen oft nur unzureichend berücksichtigt werden. Womit wir beim nächsten Punkt wären.

HERAUSFORDERUNGEN IM LOGMANAGEMENT

Um Schwachpunkte bei der Umsetzung zu vermeiden, sollten Sie die besonderen Herausforderungen im Logmanagement genau kennen. Sie sind die Angelpunkte, an denen eine Lösung ansetzen muss. Die folgenden Punkte sind hier von zentraler Bedeutung:

- **Unterschiedliche Formate**: Logdaten sind semistrukturiert. Das Format ist von Applikation zu Applikation unterschiedlich, es gibt keine allgemeingültigen Formate. Um Analyse und Korrelation zu unterstützen, müssen wichtige Logdaten also zunächst in ein einheitliches Logformat gebracht werden. Das betrifft zum Beispiel Zeitstempel.
- **Logdaten sind verteilt**: Logdaten sind über unterschiedliche Knoten verteilt. Frontend, Middleware und Backend sind nur einige Schichten in modernen Systemen, in denen Logdaten generiert werden. In jeder Schicht gibt es dann zur Lastverteilung und Ausfallsicherheit mehrere Knoten. Alle diese Daten müssen abgegriffen werden.
- **Logdatenquellen ändern sich kontinuierlich**: Gerade im Bereich Cloud-Infrastrukturen kommen neue Knoten hinzu und alte verschwinden wieder. Eine Logmanagementsoftware kann in diesem Szenario dabei helfen, Lastspitzen zu kompensieren. Dazu ist es wichtig, die entsprechenden Einsatzweisen zu kennen.

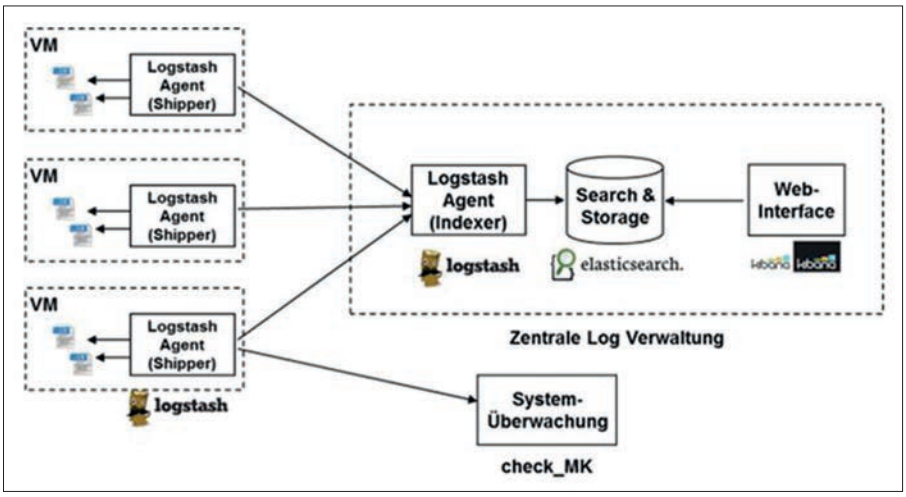

Abb. 1: Übersichtsarchitektur am Beispiel der Managementsoftware Logstash

- **Kibana:** Kibana ist eine webbasierte Anwendung. Sie ermöglicht den Zugriff auf die indizierten Daten von Elasticsearch. Eine Vielzahl von Visualisierungskomponenten lässt sich zu individuellen Dashboards zusammenstellen. Dank ihrer mächtigen Suchsyntax kann Kibana relevante Informationen in kürzester Zeit extrahieren. Diese Suchanfragen können mithilfe aktiver Checks aus anderen Monitoringlösungen zu Alarmierungszwecken eingesetzt werden.

- **Sensitive Informationen:** Logdaten können sensitive Daten enthalten, wie Nutzerkennungen und Passwörter oder andere schützenswerte Daten. Wenn ein größerer Kreis von Personen Zugriff auf Logdaten bekommt, können sensitive Informationen gefiltert werden. Eine Speicherung ist somit nicht mehr nötig.

LOGMANAGEMENTLÖSUNGEN

Logmanagementlösungen helfen bei der Bewältigung der Herausforderungen (**Abb. 1**). Im Wesentlichen bieten alle Applikationen diese Funktionen an:

- Extraktion der Logdaten
- Shipment
- Normalization und Indexing
- Suchen und Korrelieren über ein Nutzerinterface

ELK-STACK (ELASTICSEARCH, LOGSTASH, KIBANA)

ELK ist ein Open-Source-Stack aus verschiedenen Komponenten, mit dem Unternehmen ein zentrales Logmanagement aufsetzen können. Der Stack ist die Voraussetzung für weitere Anwendungen. Viele Komponenten und Konzepte des Stacks finden sich auch in anderen Lösungen wieder.

- **Elasticsearch:** Elasticsearch ist ein sehr mächtiger Volltextsuchserver, der von namhaften Internetgrößen wie XING oder Wikipedia für die Volltextsuche eingesetzt wird.
- **Logstash:** Logstash ist eine Eventpipeline, die in der Lage ist, unterschiedlichste Datenquellen, z. B. Logdateien, zu konsumieren und aufzuarbeiten. Sie extrahiert informationsrelevante Felder und reicht sie an den Suchserver Elasticsearch weiter.

Neben dem ELK-Stack erfreuen sich auch Lösungen wie die Open-Source-Software Graylog [3] und die kommerzielle Software Splunk [4] wachsender Beliebtheit. Splunk bietet als kommerzielle Lösung viele Angebote, die über das reine, zentrale Logmanagement hinausgehen. Sie sind spezifisch auf unterschiedliche Bedürfnisse und Stakeholder zugeschnitten.

LOGMANAGEMENT IN DER PRAXIS

Das folgende Beispiel soll zeigen, wie die Einführung und Umsetzung von Logmanagement auf Basis des ELK-Stacks im konkreten Projekt aussehen kann.

Der Fall: Ein Kunde führt ein großes Softwareentwicklungsprojekt durch. Mit der Software soll in der Endausbaustufe das komplette Kerngeschäft abgebildet werden. An dem Projekt sind mehrere weltweit verteilte Teams mit insgesamt über 200 IT-Experten beteiligt. In den Produktions-, Test- und Entwicklungsumgebungen werden gleichzeitig über hundert Anwendungsinstallationen, teilweise zwei bis vier Node-Cluster, bereitgehalten, die für das Testen von verschiedenen Versionsständen und bestimmten Anwendungsbereichen verwendet werden.

Mit der Einführung eines zentralen Logmanagements verfolgt das Unternehmen zwei gleichermaßen wichtige Ziele:

- Service-Desk und 2nd-Level-Support sollen in die Lage versetzt werden, mit wenigen Klicks auf den vorbereiteten Dashboards die Anwendermeldungen in den Logs nachzuvollziehen und dann gegebenenfalls den gemeldeten Fehler an das richtige Entwicklungsteam zu routen.
- Für über hundert Test- und Entwicklungsumgebungen, in denen unterschiedliche Versions- und Daten-

bestände vorgehalten werden, sollen zusätzlich zu den vorhandenen Test- und Analysewerkzeugen Möglichkeiten zum schnellen Durchsuchen, Visualisieren und Korrelieren von Logausgaben geschaffen werden.

Anforderungen des Kunden: Die Anzahl der relevanten Logdateien liegt zwischen 500 und 600. Der Kunde legt großen Wert auf ein umfassendes Monitoring und hat bereits in eine hochwertige Application-Performance-Management-Lösung investiert. Trotzdem ergeben sich immer wieder Situationen, in denen die Problem- und Fehleranalysen nur unter Berücksichtigung korrelierter Loginformationen gelöst werden können. Aus dieser Erfahrung heraus – und um künftige Analysephasen zu verkürzen – beschließt man, ein Logmanagementsystem einzurichten. Folgende Anforderungen werden für das System formuliert: Das Logmanagementsystem soll in der Lage sein,

- eine beliebige Zahl von Logs zu verarbeiten,
- alle Arten von Logdateien und Formaten zu interpretieren,
- die Visualisierung der Informationen zu ermöglichen,
- Vergleichbarkeit und Korrelierbarkeit der Informationen zu ermöglichen,
- und es soll nach Möglichkeit keine weiteren Lizenzkosten verursachen.

Die Lösung: Nach einer kurzen Recherchephase führten wir auf Basis des oben beschriebenen Stacks ein Proof of Concept (PoC) durch. Die PoC-Installation überführten wir nach kurzer Zeit in den Dauerbetrieb, wo der Kunde sie heute zur Analyse aller anfallenden Logdateien einsetzt.

Nach der Einweisung in das Logmanagementsystem können Service-Desk und 2nd-Level-Support heute spezielle Dashboards als Analysequellen für die Bewertung und die Kategorisierung von Nutzermeldungen verwenden. Damit konnte der Kunde die Qualität der Meldungen an die Entwicklung erheblich verbessern. In der Softwareentwicklung und im Testmanagement werden die Erkenntnisse, die mit der Logmanagementlösung gewonnen werden, für Fehleranalysen und zur Feinjustierung der Logging-Komponenten eingesetzt.

Der konkrete Nutzen: Die Bereitstellung eines zentralen Logmanagements trug dazu bei, die Zahl von Doppelmeldungen aus dem Service-Desk zu reduzieren. Auch die Anzahl der Anwendermeldungen, die aus dem Service-Desk 1:1 an die Entwicklungsabteilung durchgereicht werden, und die Anzahl der Meldungen, die initial an die falsche Entwicklungsabteilung geroutet werden, konnten reduziert werden.

Die Test- und Entwicklungsteams beziehen die Statistiken und Werte aus dem Logmanagement in die Auswertung von Testergebnissen und die Beurteilung der Codequalität ein.

FAZIT

Nicht nur Fachbereiche möchten wissen, wie die Geschäfte laufen und setzen dazu auf Business-Intelligence-Werkzeuge. Der Motor, der dieses Geschäft antreibt, ist für viele Unternehmen ihre IT-Anwendungslandschaft.

Zu wissen, wie dieser Motor läuft und schnelle Fehlerdiagnosen durchzuführen, ist für die Unternehmen deshalb essenziell und wichtig. Logs sind eine wichtige Quelle für diese Informationen. Bei dem Kunden aus unserem Fallbeispiel konnten Aufgaben dank dieser Informationen optimaler geroutet werden.

Einerseits ist eine „Datenzentrale“ heute wichtiger denn je. Andererseits ist aber auch der Einstieg ins Logmanagement sehr viel einfacher geworden. Für Betreiber von verteilten oder Cloud-basierten Anwendungslandschaften ist daher jetzt ein guter Zeitpunkt, um Logmanagement als wichtigen Baustein zu etablieren.

Links & Literatur

[1] Bitkom, Nutzung von Cloud Computing in Unternehmen wächst, 2014: http://goo.gl/UXoltq

[2] Humble, Farley: „Continuous Delivery: Reliable Software Releases Through Build, Test, and Deployment Automation“, Addison-Wesley Signature, Addison-Wesley, 2010

[3] Greylog: http://www.graylog.org

[4] Splunk: http://de.splunk.com/

Richard Attermeyer

arbeitet als Senior Solution Architect bei der OPITZ CONSULTING Deutschland GmbH. Er ist seit vielen Jahren als Entwickler, Architekt und Coach für die Themen Java EE, agile Projekte und Continuous Delivery tätig und hilft Unternehmen, mit motivierten Teams erfolgreiche Projekte zu realisieren.

Waldemar Kaus

ist als Solution Architect bei der OPITZ CONSULTING Deutschland GmbH tätig. Zu seinen Arbeitsschwerpunkten gehören Performanceanalysen in komplexen IT-Landschaften und die Einführung maßgeschneiderter APM-Lösungen, bei denen das Logmanagement eine feste Rolle einnimmt.