LU12a - Grundlagen der Überwachung

Die Zuverlässigkeit und Leistungsfähigkeit dieser verteilten Systeme sind von entscheidender Bedeutung für den Geschäftserfolg. Ein Ausfall oder eine Beeinträchtigung kann nicht nur zu Umsatzeinbussen führen, sondern auch das Vertrauen der Kunden beeinträchtigen und das Image einer Marke schädigen. Aus diesem Grund ist die Überwachung von verteilten Systemen ein unverzichtbarer Bestandteil des IT-Betriebsmanagements.

In der heutigen hochgradig vernetzten und digitalisierten Welt sind verteilte Systeme zu einem Eckpfeiler vieler Organisationen geworden. Diese Systeme, die aus einer Vielzahl von verbundenen Komponenten bestehen, ermöglichen es Unternehmen, komplexe Aufgaben effizient zu bewältigen und Dienstleistungen auf globaler Ebene anzubieten. Beispiele für verteilte Systeme reichen von Cloud-Infrastrukturen über soziale Netzwerke bis hin zu Finanztransaktionssystemen.

Die Überwachung von verteilten Systemen bezieht sich auf den Prozess der kontinuierlichen Beobachtung und Analyse verschiedener Komponenten eines verteilten Systems, um sicherzustellen, dass sie ordnungsgemäss funktionieren und ihren beabsichtigten Zweck erfüllen. Dabei werden verschiedene Metriken wie Systemleistung, Verfügbarkeit, Auslastung von Ressourcen, Fehlerzustände und Sicherheitsbedrohungen überwacht.

Es gibt eine Vielzahl von Werkzeugen und Technologien, die zur Überwachung von verteilten Systemen eingesetzt werden können. Diese reichen von einfachen Protokollüberwachungslösungen bis hin zu komplexen, auf künstlicher Intelligenz basierenden Analysetools. Zu den gängigen Funktionen gehören das Erfassen von Logdateien, das Überwachen von Netzwerkverkehr, das Messen der Systemleistung und das Erkennen von Anomalien.

Die Vorteile einer effektiven Überwachung von verteilten Systemen sind vielfältig. Sie ermöglicht es Organisationen, potenzielle Probleme frühzeitig zu erkennen und zu beheben, die Gesamtleistung zu optimieren, die Sicherheit zu verbessern und Ausfallzeiten zu minimieren. Darüber hinaus kann eine umfassende Überwachung auch dazu beitragen, die Einhaltung von Vorschriften und Standards zu gewährleisten, indem beispielsweise Datenschutzrichtlinien eingehalten werden.

Insgesamt ist die Überwachung von verteilten Systemen ein unverzichtbarer Bestandteil des IT-Betriebsmanagements und trägt massgeblich zur Gewährleistung der Zuverlässigkeit, Leistungsfähigkeit und Sicherheit von modernen Unternehmensinfrastrukturen bei. Durch den Einsatz geeigneter Überwachungswerkzeuge und -technologien können Organisationen potenzielle Risiken minimieren und eine reibungslose Betriebsabwicklung sicherstellen.

Komponenten

(Übersetzt von https://www.loggly.com/use-cases/distributed-systems-monitoring-the-essential-guide/)

Hardware

Die Hardware eines verteilten Systems kann in zwei Kategorien unterteilt werden: physisch und virtuell. Zur physischen Hardware gehören Server, Speicher- und Netzwerkgeräte. Virtuelle Hardware ist Software, die physische Hardware simuliert, einschließlich virtueller Maschinen und Cloud-Instanzen.

Die physische und virtuelle Hardware eines verteilten Systems lässt sich weiter in die folgenden Kategorien unterteilen:

- **Server:** Das Rückgrat eines verteilten Systems, das für die Ausführung der Anwendungen verantwortlich ist.
- **Speicher:** Der Primärspeicher speichert betriebliche Daten, während der Sekundärspeicher weniger häufig genutzte Daten speichert.
- **Netzwerk:** Switches, Router und Firewalls, die Server, Speicher und Cloud-Instanzen miteinander verbinden.
- **Cloud-Instanzen:** Virtuelle Maschinen oder Container, die in der Cloud laufen und zur Erstellung eines verteilten Systems verwendet werden.

Software

Die Software eines verteilten Systems kann in die folgenden Kategorien unterteilt werden:

- **Betriebssystem:** Die Software, die die Hardware verwaltet und eine Plattform für die Ausführung von Anwendungen bietet.
- **Anwendung:** Die Software, die auf dem Betriebssystem läuft und die Funktionen des Systems bereitstellt.
- **Datenbanken:** Eine Sammlung von Daten, auf die Anwendungen zugreifen können. Sie können in zwei Kategorien unterteilt werden: relationale und NoSQL-Datenbanken. Relationale Datenbanken sind traditionelle Datenbanken, die eine Tabellenstruktur zur Speicherung von Daten verwenden. NoSQL-Datenbanken sind neuere Datenbanken, die eine Vielzahl von Datenstrukturen, wie z. B. Schlüssel-Wert-Paare, zum Speichern von Daten verwenden.

Metriken

(Quelle: https://thwack.solarwinds.com/groups/devops/b/blog/posts/the-four-golden-signals-for-monitoring-distributed-systems)

Latenz

Als Latenz bezeichnet man die Zeit von Senden einer Anfrage (Request) bis zum Erhalten der Antwort (Response). Welche Latenz akzeptable ist, hängt von der Art der Kommunikation ab. Nachrichten zwischen zwei Applikationen müssen in der Regel eine kürzere Latenzzeit haben (Millisekunden), als die Verarbeitung einer Benutzeraktion (Sekunden).

Verkehr

Der Datenverkehr ist ein Mass für die Anzahl der Anfragen, die über das Netz laufen. Dabei kann es sich um HTTP-Anfragen an ihren Webserver oder ihre API oder um Nachrichten handeln, die an eine Verarbeitungswarteschlange gesendet werden. Zeiten mit hohem Verkehrsaufkommen können zu einer zusätzlichen Belastung Ihrer Infrastruktur führen und sie an ihre Grenzen bringen, was nachgelagerte Effekte auslösen kann. Dies ist ein wichtiges Signal, da es Ihnen hilft, Kapazitätsprobleme von unsachgemäßen Systemkonfigurationen zu unterscheiden, die auch bei geringem Datenverkehr Probleme verursachen können. Bei verteilten Systemen kann es Ihnen auch helfen, die Kapazität vorauszuplanen, um den anstehenden Bedarf zu decken.

Fehler

Fehler können Sie auf Fehlkonfigurationen in Ihrer Infrastruktur, Fehler in Ihrem Anwendungscode oder nicht funktionierende Abhängigkeiten hinweisen. Ein Anstieg der Fehlerrate könnte beispielsweise auf den Ausfall einer Datenbank oder eines Netzwerks hindeuten. Nach einer Codebereitstellung könnte dies auf Fehler im Code hinweisen, die die Tests irgendwie überlebt haben oder erst in Ihrer Produktionsumgebung auftauchen. Die Fehlermeldung gibt Ihnen weitere Informationen über das genaue Problem. Fehler können sich auch auf die anderen Metriken auswirken, indem sie die Latenzzeit künstlich herabsetzen oder wiederholte Wiederholungen verursachen, die andere verteilte Systeme überlasten.

Sättigung

Die Sättigung definiert die Belastung Ihrer Netz- und Serverressourcen. Jede Ressource hat einen Grenzwert, ab dem die Leistung abnimmt oder nicht mehr verfügbar ist. Dies gilt für Ressourcen wie CPU-Auslastung, Speichernutzung, Festplattenkapazität und Operationen pro Sekunde. Man muss das Design Ihres verteilten Systems verstehen und Erfahrung haben, um zu wissen, welche Teile Ihres Dienstes zuerst gesättigt werden könnten. Oft sind diese Messwerte Frühindikatoren, so dass Sie die Kapazität anpassen können, bevor die Leistung nachlässt.

Das Erreichen der Sättigungsgrenze kann sich auf verschiedene Weise auf Ihren Dienst auswirken. Wenn beispielsweise die CPU voll ist, kann dies zu verzögerten Antworten führen, voller Speicherplatz kann zu Fehlern bei Schreibvorgängen auf der Festplatte führen, und eine Netzwerksättigung kann zu Paketverlusten führen. Die Dashboards und Warnmeldungen von SolarWinds Observability helfen Ihnen, diese Ressourcen im Auge zu behalten und die Kapazität proaktiv anzupassen, bevor sie gesättigt werden.

M321-LU12