Planung und Erstellung einer Backend-Microservices-Architektur aus den Anforderungen durch das Spiel Stirnraten.

Michael Rothkegel

11. Mai 2019

Inhaltsverzeichnis

1 Glossar

Das ist der:
enter

Begriff	Erklärung
Deployen	Ein Softwareprodukt updaten.
Hosten	Ein Softwareprodukt auf einem Server bereitstellen. Dafür stehen heutzutage viele Möglichkeiten bereit, vom eigenen Server, gemieteten Servern oder Cloudlösungen (Azure, AWS, Google Cloud)
CI/CD	Continuous Integration/Continuous Delivery beschreibt den kontinuierlichen Prozess ein Softwareprodukt von der Entwicklung bis zur tatsächlichen Veröffentlichung zu begleiten. Dies sollte automatisch funktionieren, schnell gehen und leicht auszuführen sein.
User Interface	Oberfläche für Benutzer
VM	Virtuelle Maschine, Kapselung eines Rechnersystems innerhalb eines anderen
Legacy	Es handelt sich um Altsysteme, die eine Erneuerung benötigen oder abgelöst werden müssen. Das kann verschiedene Gründe haben, wie z.B. schlechte Programmierung, Fehlentscheidungen oder technologische Veraltung
Technologiestack	
Overhead	
Downtime	Zeitpunkt, die eine Anwendung benötigt, bis sie neugestartet ist. Während der Downtime ist eine Anwendung nicht erreichbar.
technische Schuld	
Monolith	monolitsche Struktur
Microservice	Service
Domain Driven Design (DDD)	
Bounded Contexts	
up-stream	
down-stream	
Pattern	
REST	
Monitoring	
Deployment	
Logging	
Traffic	
Timeout	
Fallback	
Route	

Eventbus
Subscriber/subscribe
publish
Message Broker
Load Balancer
REST
Backend
Performance
Latenz

Tabelle 1: Glossar

2 Einleitung

3 Grundlagen

Hier erklären was kommt

3.1 Microservices

Für den Begriff Microservices existiert keine einheitlich anerkannte Definition. Während Wolff unter Microservices unabhängig, deploybare Module versteht[13], spricht Newman von kleinen, autonomen Services, die zusammenarbeiten. Cockcroft verwendet den Begriff Microservice gekoppelt mit einem Architekturbegriff: Eine Microservice Architektur sind gekoppelte Services, welche für einen gewissen Kontextbereich zuständig sind.[4] D.h. jeder Service behandelte gewisse, fachliche Aufgaben und kann genau für diese genutzt werden. Eine Vielzahl von solchen Services bildet dann die gesamte Anwendung.

Amudsen schreibt dem Microservice an sich die Eigenschaft zu, dass er unabhängig zu anderen Microservices sein muss, d.h. ein Microservice kann losgelöst von anderen geupdated (deployed) werden. Weiter ist ein Microservice wie schon bei Cockcroft für einen gewissen Aufgabenbereich zuständig. Eine Microservice-Architektur ist ein zusammenschluss von miteinander kommunizierenden Microservices.[4]

In *Flexible Software Architecture*[14] werden Microservices zu den bisherigen noch weitere, teils technische Eigenschaften zugeschrieben: Microservices sind technologisch unabhängig, d.h. eine Microservice Architektur ist beispielsweise nicht an eine bestimmte Programmiersprache oder Datenbank gebunden. Weiter müssen Microservices einen privaten Datenspeicher haben und sie kommunizieren mit anderen Services über das Netzwerk (z.B. über REST). Ebenfalls werden Microservices verwendet, um große Programme in kleine Teile zu unterteilen. Diese kleine Teile lassen sich automatisch bauen und deployen.

Basierend auf den folgenden Definitionen wird der Microservice Begriff wie folgt verwendet: Microservices sind

- klein in der Größe
- kommunizieren mit anderen Services über Netzwerkschnittstellen (z.B. REST) sind unabhängig voneinander deploybar
- können unabhängig voneinander entwickelt werden (d.h. Microservice A muss nicht auf B,C,D ... warten und/oder umgekehrt)
- eingeschränkt in ihrer Geschäftslogik, d.h. ein Microservice kümmert sich immer um einen speziellen Kontext, der im vorhinein definiert werden muss
- dezentral, d.h. sie können auf unterschiedlichsten Plattformen gehosted werden werden automatisch gebaut und deployed

Abschließend handelt es sich um eine Microservice-Architektur, wenn viele Microservices nach Definition verwendet werden.

3.2 Monolithische Struktur

Eine monolithische Struktur ist ein einziges Softwareprogramm (Monolith), welches in sich geschlossen ist. Dies bedeutet im Detail, dass ein Monolith aus mehreren Ebenen besteht auf die über Schnittstellen zugegriffen werden kann. Innerhalb der Ebenen werden Komponenten wie z.B. Frameworks oder selbstgeschriebene Klassen eingebunden und verwendet.[1] Durch die sich aufeinander aufbauenden Ebenen folgt daraus, dass sämtliche Geschäftslogiken, User Interfaces sowie die Datenbank und Datenbankzugriffe Abhängigkeiten haben. All dies ist in einem Programm vereint.[1] Natürlich kann die monolithische Struktur innerhalb noch einmal Modular sein. In einem Monolithen existierten also ggf. mehrere Module, welche verschiedene Geschäftslogiken abbilden oder ein Modul, welches nur zum Erstellen einer grafischen Oberfläche verwendet wird. Dennoch können diese Module nicht unabhängig von der gesamten Anwendung deployed werden.[6]

Abbildung 1 zeigt eine vereinfachte Gegenüberstellung der beiden Architekturen. Die App 1 ist in drei klassische Funktionen (Web, Business und Data) unterteilt. Die Skalierung (2) kann durchgeführt werden, in dem App 1 über mehrere Server oder VMs geklont wird.

Bei der Microservice Architektur werden die Funktionen auf unterschiedliche Dienste aufgeteilt. Konkreter könnte dies bedeuten, dass App 1 bei (3) zuständig für eine Benutzerkontoverwaltung ist und App 2 für ein Abrechnungssystem. Die Microservices (4) werden nicht geklont, sondern können unabhängig voneinander bereitgestellt werden.

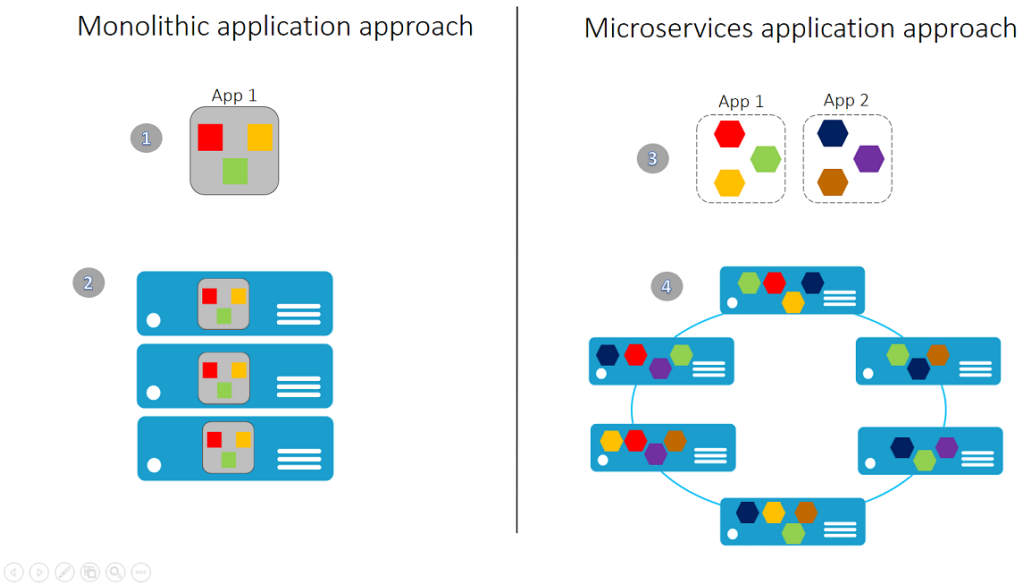

Abbildung 1: Monolith und Microservice-Architektur.[1]

3.3 Monolith vs. Microservices

Aus den vorherigen Abschnitten sind diverse Unterschiede zwischen den Architekturen erkennbar. Nun gilt es festzustellen, für welche Problemstellungen, welche Architektur sinnvoller ist. [14] [2]

Aus der Tabelle ergeben sich verschiedene Punkte: Der Monolith eignet sich besonders dann sehr gut, wenn die Projekt- sowie Teamgrößen absehbar sind und auch die Technologie entschieden ist. Zusätzlich kann es beim Projektanfang ein Vorteil sein, da die Abhängigkeiten innerhalb des Projektes liegen und so Entwicklungsgeschwindigkeit nicht durch komplizierte Infrastrukturen blockiert wird. Ist die Projektgröße allerdings nicht absehbar, treten früher oder später mehrere Schwierigkeiten auf: Zum einen bindet der am Anfang des Projektes festgelegte Technologiestack und die Nutzung oder der Austausch neuer Technologien sind in der Regel mit sehr viel Arbeit verbunden. Des Weiteren führen die anfangs eingegangen Abhängigkeiten zu problemenen im Deployment (A kann erst updaten, wenn B soweit ist) und einem erhöhten Aufwand in der Kommunikation zwischen den Teams (A kann erst beginnen, wenn B xy erledigt hat).

Zusätzlich ist die Skalierung von Microservices unabhängiger. Es können sich feingranular Services gesucht werden, welche skaliert werden sollen. Diese benötigen nicht zwingend mehr Hardware (vertikale Skalierung), sondern könnten z.B. auch auf verschiedene Server verteilt werden (horizontale Skalierung). Dies ist bei einem Monolithen natürlich auch möglich, dennoch muss immer der ganze Monolith skaliert werden, welcher zum einen immer mehr Hardware als einzelne Microservices benötigt und zum anderen auch durch die Komplexität in der Regel auch schwerer zu skalieren ist.[13] Als abschließender Punkt ist die Robustheit zu erwähnen: Wenn ein Microservice einen Fehler enthält,

	Monolithische Architektur	Microservice-Architektur
Abhängigkeiten	alles in einer Anwendung	entkoppelt, da Prinzip von Modularisierung verwendet wird
Größe	linear steigend	einzelne Services sind klein
Geschwindigkeit Zugriffe	schnell, da alles in einer Anwendung	Zugriffe können länger dauern
Deployment	schwieriger desto größer das Projekt, aufgrund von • Abhängigkeiten • Größe	einfach, da Microservices • klein und • modular sind
Organisation	leichter, da alles an einem Ort	schwerer, da mehr Domänenlogik (wer macht was?) beachtet werden muss
Legacy-Systeme ablösen	ggf. schwierig, da System sehr verzahnt miteinander	leicht, da Microservices durch neue abgelöst werden können
Technologie	beschränkt	vielfältig
Nachhaltige Entwicklung	wartbar mit Einschränkungen	leicht wartbar
Robustheit	weniger, da ganzes System bei schweren Fehlern abstürzt	sehr, da im Zweifel immer nur ein Service abstürzt
Skalierung	horizontale und vertikale Skalierung, Umsetzung kann sehr komplex werden	horizontale und vertikale Skalierung
Betrieb	nur ein System	komplex, da mehr Services verwaltet werden müssen

Tabelle 2: Monolitih vs. Mircoservice-Architektur

stürzt dieser im schlechtesten Fall ab. Im besten Fall übernimmt dieser Service eine weniger wichtige Funktion und der Nutzer bemerkt den Ausfall noch nicht einmal. Beim Monolithen dagegen stürzt die gesamte Anwendungen ab. In der Regel startet so eine Anwendung automatisch neu, jedoch ist betrifft die Downtime alle Nutzer.

Aus den genannten Punkten lässt sich schließen, dass eine generelle Aussage, ob eine monolithische oder Microservice-Architektur besser oder schlechter ist, sich nicht treffen lässt. Es kommt immer drauf an, welche Zielsetzung und wie viele Ressourcen für das Projekt festgelegt sind. REWE Digital beispielsweise hat ihr Produkt zuerst als Monolithen gestartet und ist erst später auf eine Microservice-Architektur umgeschwenkt.[3] Zwei mögliche Gründe könnten dafür sein, dass zum einen ein lauffähiges Produkt schneller mit einer monolithischer Struktur zu erreichen ist, zum anderen ist nicht gegeben, ob ein entwickeltes Projekt überhaupt die Nachfrage erzeugt, so dass eine Microservice-Architektur notwendig ist. Dementsprechend muss abgewogen werden, welche Architektur für welchen Anwendungsfall besser geeignet ist.[13]

3.4 Architektur von Micrsoservices

Wie bereits erwähnt, ist die Entkopplung von Microservices ein großer Vorteil gegenüber dem Monolithen. Dennoch ist es sinnvoll Richtlinien, Regeln und/oder Festlegungen zu schaffen, damit die Microservices nicht blockierend oder technologisch unnötig gegensätzlich arbeiten. Die Entscheidungsebene kann global (Makroarchitektur) oder nur für einen einzelnen Services (Mikroarchitektur) gelten.[14] Welche Festlegungen und mit welcher Strenge diese eingehalten werden müssen, hängt von verschiedenen Faktoren ab, welche technologisch, organisatorisch oder wirtschaftlich motiviert sein können.[3]

In dem folgenden Abschnitt wird das Grundprinzip der Software-Modellierungs-Methodik Domain Driven Design untersucht. Zusätzlich wird erläutert, welche Fälle makro- oder microarchitektonisch einzuordnen sind.

3.4.1 Domain Driven Design

Domain Driven Design (DDD) ist ein Vorgehen mit dem ein Softwaresystem modelliert werden kann. Im Sinne einer Microservice-Architektur kann dies als Werkzeug genutzt werden, um Microservices fachlich einzuteilen.[9] Beim sogenannten *Strategic Design* wird dafür das Softwaressytem in verschiedene *Bounded Contexts* eingeteilt, welche an ein *Domänenmodell* gebunden sind. Ein Domänenmodell bildet die Geschäftslogik ab, d.h. inwiefern einzelne Objekte innerhalb des Kontexts in Relation zueinander stehen, welche Eigenschaften sie haben und wie sich verhalten. Dabei kann ein Domänenmodell - je nach Entwurfsmuster - von einem oder mehreren Bounded Contexts genutzt werden.[13]

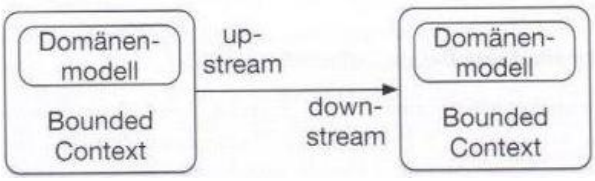

Abbildung 2: Bounded Contexts mit eigenständigem Domänenmodell.[13]

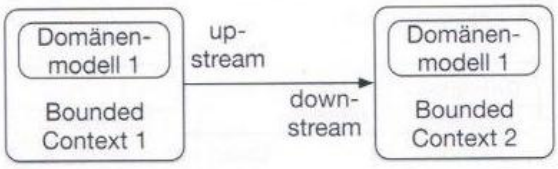

Abbildung 3: Bounded Context 2 adaptiert das Domänenmodel von Bounded Context 1.[13]

Wolff verdeutlicht das Prinzip von Bounded Contexts mit Hilfe von vier Microservices, welche einen Onlineshop repräsentieren: *Suche*, *Check Out*, *Inkasso* und *Lieferung*. Während das Datenmodell der Suche detaillierte Informationen über die Produkte enthält, reicht es im Warenkorb (hier Check Out), wenn ggf. nur der Produktname gespeichert wird. Bei Inkasso ist es ähnlich: An dieser Stelle sind Zahlungsdaten des Benutzers relevant, während bei Lieferung die ggf, nur Adresse notwendig ist.[13] An diesem Beispiel wird deutlich, dass zwar von selben Begrifflichkeiten wie Benutzer, Produkt usw. gesprochen wird, allerdings jeder Service sein eigenes Domänenmodell hat. Durch diese Technik, was dabei hilft, eine saubere Microservice-Architecktur zu erstellen.[13] [9]

Neben der fachlichen Trennung bildet DDD auch die Kommunikation zwischen Kontexten ab. Dabei wird grundsätzlich vom *up-stream* (vorgeschaltet) und dem *down-stream* (nachgeschaltet) gesprochen.[13]. Der up-stream stellt dem down-stream Informationen bereit. Wie dies technisch umgesetzt ist, also ob der down-stream nachfragt oder der up-stream aktiv Daten schickt, ist frei wählbar.

Nach dem Anwenden von DDD sollte die Struktur der Software erkennbar sein: D.h. welche Art Microservices werden benötigt und iwiefern sie mit anderen in Abhängigkeiten bzw. Kommunikation stehen.

3.4.2 Macro- und Mikroarchitektur

Wenn durch das DDD entworfen wird, welche Microservices voraussichtlich benötigt werden, ist es sinnvoll einen Art Bauplan zu verfassen, welcher impliziert, an welche Regeln sich ein Microservice halten muss. Diese Regeln können wie bereits erwähnt auf globaler Ebene getroffen werden, d.h. sie gelten für alle Services (Makroarchitektur) oder sie gelten nur im Microservice selber (Microarchitektur). REWE Digital unterscheidet dabei zwischen *Must*, *Should*, *Could*. D.h. es gibt Regeln, die Microservices erfüllen müssen wie

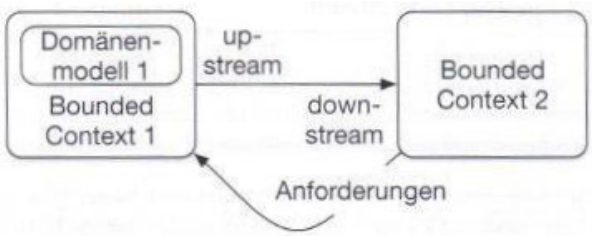

Abbildung 4: Bounded Context 2 erhält ein auf ihn zugeschnittenes Domänenmodel von Bounded Context 1.[13]

z.B. das Kommunzieren über REST oder das Implementieren eine einheitlicher Autorisierung.[3] Andere Regeln dagegen sind viel mehr Richtlinien (should) oder komplett optional (could). Das Ziel ist stets, dass durch die makroarchitektonischen Entscheidungen nicht die Vorteile von Microservices beschnitten werden.[13][4].

Es gibt verschiedene Einflussfaktoren wie sich die Makroarchitektur für ein Unternehmen definiert: Zum einen empfiehlt es sich ein Gremium zu gründen, welches sich stetig mit den Regeln der Makroarchitektur auseinandersetzt, sie entsprechend erweitert, überarbeitet und die getroffenen Entscheidungen auch immer begründen kann.[13] Zum anderen besteht immer ein technischer Einfluss[13]:

- Gewählte Technologien müssen in die Infrastruktur des Unternehmens passen: Angenommen die Auslastung eines Microservices muss überwacht werden und dies wird firmenweit mit Tool A erledigt, dann wäre es sehr aufwendig, wenn der besagte Service nur eine Schnittstelle für Tool B anbietet und der nächste Service nur für Tool C. Dies würde einerseits sehr unübersichtlich werden und andererseits viel Aufwand bedeuten.

- Technologien sind immer von dem Personal abhängig: Gerade wenn Unternehmen klein bis mittelständig sind, emphielt es sich Technologien zu nutzen, die mehrere Entwickler beherrschen, um Inselwissen zu reduizieren.

- Ebenfalls können gezielt strategische Entscheidungen getroffen werden, z.B. wenn ein Unternehmen die Dateninfrastruktur zu einem Cloudanbieter auslagern möchte, hat dies entsprechende makroarchitekonisches Auswirkungen.

Basierend auf Wolff und Nadareishvili ist folgende Tabelle entstanden, welche einen Überblick darüber gibt, wie gängige Entscheidungspunkte einzuordnen sind. [13][4][3]

In der Tabelle sieht man, dass gerade die ersten Punkten sehr von der Unternehmenskultur und den technischen sowie personellen Freiheiten abhängt. In der Theorie sollte die Programmiersprache sinnvoll für jeden Microservices gewählt werden, dennoch ergibt es auch Sinn einen Pool an Programmiersprachen auf Makroebene zu definieren, um Inselwissen zu reduzieren und nachhaltige Codequalität zu gewährleisten. Ähnliches gilt beispielsweise für die Wahl der Datenbank: Ist bereits eine globale Infrastruktur für Datenbank X geschaffen, sollte diese nicht ohne weiteres aufgebrochen werden, nur weil

	Mikroarchitektur	Makroarchitektur
Programmiersprache	x	x
Datenbank	x	x
Look and Feel (UI)	x	x
Dokumentation	x	x
Datenformat		x
Kommunikationsprotokoll		x
Authentifizierung		x
Integrationtests		x
Autorisierung	x	
Unittests	x	
Continuous-Delivery-Pipeline	x	

Tabelle 3: Entscheidungen Micro- und Macroarchitektur

es technisch möglich ist.

Bei der Dokumentation sowie beim Look & Feel ist es sinnvoll, globale Richtlinien zu definieren, damit klar ist, wo bei jedem Microservices die Dokumentation zu finden ist oder wie ein User Interface grundsätzlich angeordnet und gestaltet werden soll. Dennoch können diese Punkte im Detail je nach Microservice abweichen.

Ein Kommunikationsprotokoll (z.B. REST) sowie Datenformate (z.B. JSON) sollten festgeschrieben werden.[3] [13] Als Grund wird zum einen das Vermeiden von technischen Mehraufwand angegeben, zum anderen sind Microservices zwar eigenständig deploybare Einheiten, dennoch sollten sie technisch zum Gesamtsystem passen und nicht dagegen arbeiten.

Während die Authentifzierung (um wen handelt es sich) einmalig festgelegt werden sollte, liegt die Überprüfung der Autorisierung (was darf der Benutzer) in jedem Microservice selbst. Die Alternative wäre, dass jede eingehende Anfrage noch einmal geprüft wird, was zu unnötig hohem Traffic und Verzögerungen führen würde.

Die hier erarbeitete Tabelle ist an dieser Stelle nicht als feststehendes Manfifest für alle Unternehmen zu verstehen, sondern als neutral betrachtete, sinnvolle Einordnung. Natürlich können architektonische Entscheidungen stark vom jeweiligen Anwendungszweck abhängen.

Die Tabelle zeigt lediglich allgemeine Beispiele und ist nicht als vollständig zu betrachten. Nicht aufgeführt ist beispielsweise der Umgang mit Konfigurationsdateien, Monitoring oder Logging. Dies liegt unter anderem daran, weil es auf Projekte oder von dem

Microservice abhängt: Beim Monitoring könnte zum Beispiel global entschieden werden, *wo* Metriken abgelegt werden bzw. mit *welcher* Technologie gearbeitet wird. Aus microarchitektonischer Sicht könnten die Services selbst entscheiden, *was* gemessen wird.

Ebenso beim Deployment: Es gibt zahlreiche Methoden, um neue Updates bereitzustellen wie z.B. mittels Docker, Kubernetes oder indivudelle Installationsskripte.[13] Welche Technologie sich durchsetzt, muss anhand der Anforderung entschieden werden.

Aus den erarbeitenten Punkten lassen sich Vor- und Nachteile ableiten, zwischen denen abgewogen werden muss. Vorteile für microarchitektonische Entscheidungen sind ein sehr hohes Maß an Flexibilität und eine hohe Unabhängigkeit im Gesamtsystem, was grundsätzlich das Ziel von Microservices ist. Dies wiederum kann dazu führen, dass Entwicklungsoverhead oder Inselwissen entsteht. Ebenfalls könnten Punkte wie z.B. das *Look & Feel* oder die *Dokumentation* darunter leiden.

Macroarchitektonische Entscheidungen haben zum Vorteil, dass es Regeln gibt, welche die Entwicklung vereinfachen sollen und gegebenfalls die Nachteile der Microarchitektur kompensieren können. Auf der anderen Seite schränken macroarchitekonische Entscheidungen ein. Zusätzlich müssen sie organisch, z.B durch ein extra dafür geschaffenes Gremium, durchgesetzt und werden.[13] Im Gesamten lässt sich daraus schließen, dass sehr genau abgewogen werden muss, welche Entscheidungen global oder individuell entschieden werden. Grundsätzlich gilt, dass jede Entscheidung begründbar sein muss.

3.5 Kommunikation

Bei einem Monolithen wird eine Abfrage über eine Route gestellt, woraufhin die Anwendung entsprechend mit der Bearbeitung beginnt. Da die gesamte Datenhaltung an einer Stelle ist, sind alle Daten bekannt und abrufbar. Wichtiger noch: Die Daten sind konsitent.

Microservices sind diesbezüglich herausfordernder. Es müssen verschiedene architektonische Entscheidungen getroffen werden, wie z.B. ob es einen zentralen Service gibt, welcher alle Anfragen weiterleitet (**API Gateway**) oder ob jeder Service einzeln erreichbar ist. Ebenfalls sollte auch begründbar entschieden werden, ob eine synchrone, asynchrone oder möglicherweise eine hybride Kommunikation verwendet wird.

In den folgende Unterkapiteln werden Vor- und Nachteile der verschiedenen Kommunikationsarten für Microservices mit einem Schwerpunkt auf Unabhängigkeit (Entkopplung) und Datenkonsitenz untersucht.

3.5.1 Synchrone Kommunikation

Wenn ein Microservice bei der Bearbeitung einer Anfrage selbst eine weitere Anfrage an einen anderen Microservice stellen muss und auf das Ergebnis wartet, spricht man von synchroner Kommunikation.[13]

Anhand dieser Definition lässt sich folgendes Szenario darstellen (siehe Abbildung 5).

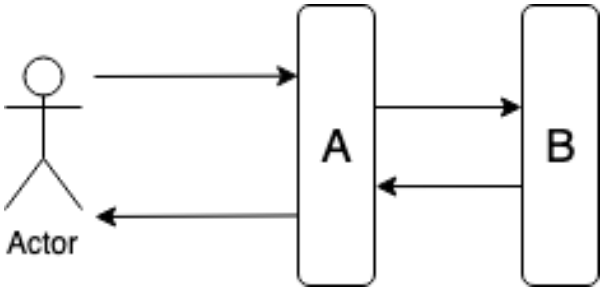

Abbildung 5: Synchrone Kommunikation

Der Actor fragt bei Microservice A an, welcher diese Anfrage bearbeitet und schließlich an B weiterleitet. B verarbeitet die Anfrage und antwortet, schließlich kann auch A antworten. Aus diesem Ablauf lässt sich festhalten, dass die übertragenden Daten aktuell sind. D.h. der Actor erhält definitiv konsistente Daten, was positiv zu vermerken ist. Problematischer dagegen ist die Abhängigkeit, welche entsteht. Sollte B nicht erreichbar sein, läuft A in einen Timeout und ist blockiert. Einerseits ließe sich argumentieren, dass genau dies passieren soll, schließlich scheint es einen Fehler zu geben. Aber angenommen A wäre ein Service zum Erstellen von Rechnungen und B ein Service zum Sammeln von Daten. A möchte B infomieren, dass eine Rechnung erstellt wurde, ist aber blockiert. Die Operation eine Rechnung zu erstellen hätte höhere Priorität als es statisch zu erfassen. Nach den aufgestellen Definitionen aus 3.3 für Microservices wird die Entkopplung, Modularität und Robustheit des Systems verletzt, da A nicht weiterarbeiten kann.[13] [5]

Bläht man das Beispiel auf, so dass weitere Services statistische Daten erfassen wollen, würden zahlreiche Microservices ausfallen (siehe Abbildung 6).

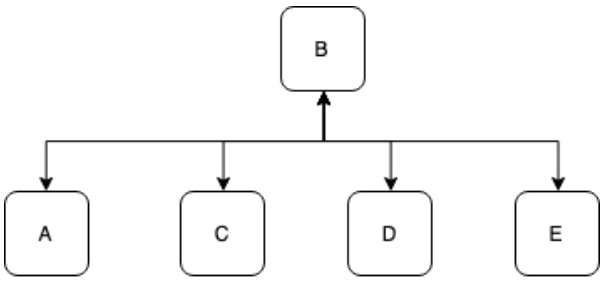

Abbildung 6: Abhängigkeiten in synchroner Kommunikation

Die Microservices können auch nicht auf eine Notfallstrategie (**Fallback**) zurückgreifen. Angenommen wenn Serivce B nach x Sekunden nicht erreichbar ist, wird die ursprüngliche Abfrage durch A weiter abgearbeitet, um nicht zu blockieren. Nun müsste eine andere Logik dafür sorgen, dass die Daten, die nicht übertragen werden konnten,

zu einem anderen Zeitpunkt übertragen werden. In diesem Moment herrscht keine Konsistenz mehr vor, was aber ein großer Vorteil an synchroner Kommunikation ist.

Betrachtet man das Beispiel andersherum und nimmt an, dass jeder Service seine Datenhaltung soweit aufspreizt, dass jeder Service die Statistiken führt, ist es nur eine Frage der Zeit, bis weitere Felder gespeichert werden müssen. Überspitzt formuliert, würde jeder Service alles speichern, was gegen die Absicht von Bounded Contexts arbeiten würde.[13]

Auch nicht zu vernachlässigen, ist die Geschwindigkeit mit der die Abfragen abgearbeitet werden können. Möglicherweise muss B in einem anderen Szenario noch mit C kommunizieren. Die Anfrage würde sich über drei Services erstrecken, was zusätzliche Latenzzeiten mit sich bringt.[13]

Ebenfalls entsteht durch jede Schnittstelle eine fachliche Abhängigkeit geschaffen.[5] Die Anfragen von den Microservices A, C, D, E müssen der Schnittstellendefinition von B entsprechen. Änderungen führen gegebenenfalls zu Fehlern und weiteren Abhängigkeiten. Wie diese Problematik gelöst werden könnte, wird 3.5.2 beschrieben.

3.5.2 Asynchrone Kommunikation

Wie bereits beschrieben, wird bei der synchronen Kommunikation auf weiterführende Abfragen gewartet. Die asynchrone Kommunikation wartet nicht auf Antworten von weiteren Services, sondern trifft Annahmen über etwaige Systemzustände.[13] Um Annahmen zu treffen, existieren je nach Anwendungsfall verschiedene Strategien:

1. Ein Microservice kann replizierte Daten vorhalten. Angenommen ein Artikel soll rausgeschickt werden: Der dafür verantwortliche Service benötigt die Anschrift des Kundens, aber nicht weitere Daten wie Geburtsdatum, Zahlungsmethode oder ähnliches. Dementsprechend werden nur relevante Daten repliziert vorgehalten. Eine Herausforderung ist es, dass diese replizierten Daten stets mit den Originaldaten übereinstimmen. Schließlich kann sich eine Anschrift ändern.[13]

2. Ggf. muss nur ein weiterer Service informiert werden wie der Service B aus Abschnitt 3.5.1, welcher Statistiken erfasst. In dem Szenario der asynchronen Kommunikation würde die Abfrage gestellt werden ohne das Ergebnis abzuwarten, da es schlichtweg nicht relevant ist. Die Herausforderung hier ist, zu gewährleisten, dass die Abfrage auch in Fehlerfehlen früher oder später zugestellt wird.

Aus den Strategien ergeben sich Anforderungen an die Kommunikationsstruktur: Es muss gewährleisten sein, dass fehlerhafte Abfragen erneut übermittelt werden und ebenfalls wird eine Struktur benötigt, die dafür sorgt, dass replizierte Datensätze stets mit aktuellen Daten befüllt sind. Dies lässt sich durch sogenannte Events erreichen.[5].[13]

Die folgende Abbildung 7 verdeutlich das Prinzip von Events und deren Infrastruktur:

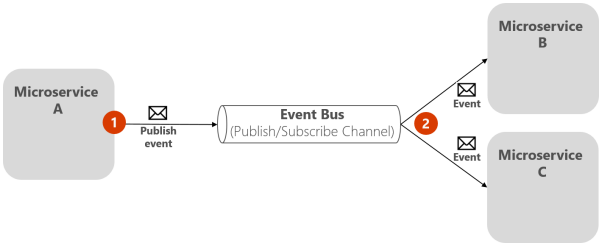

Abbildung 7: Eventbus mit Events[10]

In dieser Abbildung haben Microservice B und C das Event x beim Ergeinisbus (**Event Bus**) abonniert (**subscribed**). D.h. Microservice A veröffentlicht (**published**) eine Änderung, woraufhin B und C informiert werden. B und C können nun ihren Datenbestand aktualisieren und halten so die aktuellen Daten vor. Parallel kann der Microservice A seine Abfrage ganz normal weiterführen. Der Eventbus ist dementsprechend ein Vermittler (**Message Broker**), welcher garantiert, dass die Nachrichten übertragen werden. Dieser sollte mit etwaigen Fehlerfällen (z.B. C ist nicht erreichbar) umgehen können und eine spätere Übertragung garantieren. [10][13]

Aus diesem Modell ergibt sich ein weiterer Vorteil, nämlich dass die Microservices entkoppelt sind. Es wird keine REST-Schnittstelle definiert, welche eine gewissen Fachlogik vorgibt. Ebenfalls können mehrere Services auf ein Event hören. Wolff warnt allerdings davor Events unnötig aufgebläht zu gestalten: Zum einen werden schnell Daten übermittelt, die nicht für alle Abonnenten (**Subscriber**) relevant sind und zum anderen entspräche dies nicht dem Prinzip vom DDD.

Zusätzlich sollte beachtet werden, dass Microservices so gestaltet werden, dass sie idempotent sind. In diesem Zusammenhang bedeutet dies, dass falls ein selbes Event zweimal übertragen wird, der Microservice die Aktion nicht zweimal ausführt. D.h. eine Mehrfachausführung führt zu dem selben Ergebnis wie eine einzige Ausführung. Wenn z.B. eine Rechnung versendet werden soll, ist garantiert, dass diese nur ein einziges Mal versendet wird.[13]

3.5.3 Abwägung asynchrone vs. synchrone Kommunikation

Aus den zwei vorherigen Abschnitten ergibt sich folgende Aufstellung (siehe Tabelle 4).

	synchrone Kommunikation	asynchrone Kommunikation
Vorteile		
	• Jederzeit Konsitent • Paradigma ist Entwicklern bekannt[13]	• Entkoppelt durch Events • Flexibilität, da ein Event mehre Services erreichen kann • Nachrichtenempfang garantiert (ggf. mit Verzögerung) • Absicherung gegen Ausfall
Nachteile		
	• Anfälligkeit durch Abhängigkeiten • Erweiterbarkeit ist schwerer, da fachliche Abhängigkeiten • Ggf. lange Netzwerkzeiten	• nicht jederzeit garantiert konsistent • Idempotenz muss beachtet werden

Tabelle 4: synchrone vs. asynchrone Kommunikation

Es lässt sich feststellen, dass die Vorteile einer asynchrone Kommunikation für Microservices überwiegen und auch diese wird empfohlen.[13][5] Allerdings ist die Kommunikation nicht dogamtisch zu betrachten, sondern sollte je nach Projekt und Anwendungsfall entschieden werden. Synchrone Kommunikation bietet sich nämlich gerade dann an, wenn der Datenbestand definitiv konsistent sein sollen.

Das sogenannte CAP-Theoreom beschreibt die Abwägung, welche man in verteilten Systemen bei der Auswahl der Kommunikation treffen muss. CAP bedeutet:

- Consistency (Konsistenz): Die Daten in einem verteilten System sind konsistent.
- Availability (Verfügbarkeit): Die Verfügbarkeit für alle Systeme ist gegeben.
- Partition Tolerance (Partitionstoleranz): Das Gesamtsystem arbeitet auch weiter, wenn Teile davon ausfallen.

In einem verteilten System können immer nur zwei von den drei Bedinungen erfüllt sein.[13]. Sofern Konsistenz gewährleistet soll, müssen alle dienste stets verfügbar sein. Damit kann der Punkt Partitionstoleranz nicht erfüllt sein. Umgekehrt: Wenn die Partitionstoleranz garantiert ist, z.B. dadurch dass Services ihre eigene Datenhaltung besitzen, ist zwar prinzipiell auch die Verfügbarkeit gegeben, aber nicht die Konsistenz.

Dementsprechend ist es sinnvoll sich die Anforderungen, welches man an sein System hat zu überlegen und sich aufgrund dieser Grundlage zu entscheiden, welche Kommunikationsart implementiert werden soll.

3.5.4 API-Gateway

Umso mehr Services aufgesetzt werden, desto komplexer ist es, die Übersicht über alle zu behalten. Eine Abhilfe im Routing bietet ein sogeanntes API-Gateway. Ein API-Gateway ist der einzige Einstiegspunkt für den Nutzer (**Client**). Von dort wird er weitergeleitet, ohne die Routen von einzelnen Services zu kennen.

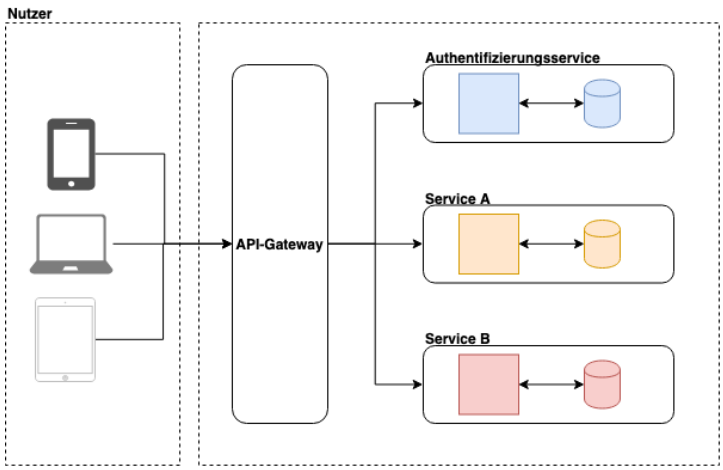

Abbildung 8: Prinzip API-Gateway

Abbildung 8 zeigt deutlich wie die Clients über das Gateway kommunizieren, welches anschließend an die entsprechenden Services weiterleitet. Wenn APIs eine hohe Auslastung haben, würde man mehrere Instanzen von einem API-Gateway erstellen. Ein sogenannter Load Balancer wäre der Einstiegspunkt für die Clients. Dieser würde die Anfragen sinnvoll an die API-Gateway-Instanzen verteilen, so dass keine Überlastung entsteht.[11]

API-Gateways haben neben dem einzelnen Einstiegspunkt noch weitere Vorteile:

- Höhere Sicherheit, das einzelne Services nicht sichtbar und nur über das Gateway zu erreichen[5]
- Authentifizierung kann bereits im Gateway ausgeführt werden, dies führt zu weniger Last für einzelne Microservices.[13]

- Zentralisiertes Logging, Caching, Monitoring, Mocking sowie eine zentralisierte Dokumentation ist möglich.[13]

Ein Nachtteil in der Struktur des API-Gateways ist, dass die Abfragen länger sind, da sie immer erst über das Gateway gehen.

3.6 Authentifizierung und Autorisierung

Beim Monolithen ist architektonisch klar, dass die Authentifizierung und Autorisierung innerhalb des Monolithen stattfindet. Im Bereich der Microservice Architektur existieren verschiedene Szenarien, wie man eine Authentifizierung sowie Autorisierung gestalten kann.[5]

Authentifizierung: Identifiziert, wer jemand ist. Z.B. Nutzer A, der sich durch Benutzername und Passwort registriert hat.[5] **Autorisierung**: Bestimmt, wie viel ein Nutzer darf. Nutzer A hat eine Rolle, welche ihn berechtigt gewisse Aktionen durchzuführen.[5]
Es emphielt sich die Autorisierung in den Microservices ansich zu überprüfen, da diese den entsprechenden Datenbestand haben. Um unnötige Last zu verhindern, kann die Validität - ob es sich überhaupt um einen gültigen Request handelt - bereits im API-Gateway überprüft werden. Die Authentifizierung dagegen sollte in einem eigenen Service oder ins API-Gateway verlagert werden.[3][8] Zu empfehlen ist, dass die Authentifizierung an einer zentralen Stelle durchgeführt wird, um Redundanz und fehlerhafte Implementierungen zu verhindern.

Die Idee ist, dass der Benutzer nach dem Anmelden ein Security-Token erhält, welches verwendet wird, um sensible Anfragen zu verifizieren. Zum einen gibt es die Möglichkeit ein Token auszustellen, welches beim Auslesen verschlüsselt ist (opaque Token) und zum anderen auf ein offenen Standard namens Json Web Token (JWT, transparent Token) zu setzen. Das opaque Token hat den großen Nachteil, dass es zusätzliche Performance sowie Latenz verursacht und nur synchron entschlüsselt werden kann.[8]

Das JWT wird beim Ausstellen signiert, um die Echtheit zu gewährleisten. Während die Nachteile des opaque Token hier nicht auftreten, ist ein anderes Problem, dass ein JWT nach Ausstellung nicht widerrufen werden kann. Theoretisch wäre es dauerhaft gültig, weshalb Ablauflaufzeiten gesetzt werden. Dies wiederum impliziert, dass der Client dafür sorgen muss, immer rechtzeitig ein neues Token anzuforden. Für solche und weitere Logiken exisitiert bereits ein Sicherheitsstandard namens OAuth2, welcher empfohlen wird zu verwenden.[8]

Ziel bei OAuth2 ist unter anderem Autorisierungen zwischen verschiedene Anwendungen zu erlauben. Ursprünglich wurde das Authentifizierungsprotokoll so entworfen, dass Drittanwendungen Zugang zu Informationen erhalten, ohne dass Passwörter weitergelei-

tet werden müssen.[8] Beispielsweise wird OAuth2 verwendet, wenn Benutzer sich über ihren Facebook-Account bei Drittplattformen anmelden. Die Drittplattformen können natürlich nicht das Facebook-Passwort einsehen, erhalten aber je nach Anwendungsfall Zugriff auf verschiedene Ressourcen (z.B. Lesezugriff auf die E-Mail-Adresse und/oder Konakte, Schreibzugriffe zum Teilen von Nachrichten usw.).

Da OAuth2 ein sehr komplexes und umfangreiches Thema ist, wird im Folgenden nur ein häufig verwendetes Grundprinzip erklärt.

Um die Abbildung 9 besser zu verstehen, sind folgende Definitionen hilfreich:[8]

Authorization Server: Authentifiziert den Benutzer und gibt ein Access sowie Refresh Token raus.

Access Token: Durch ein Access Token erhält man Zugriff auf den Resource Server. Das Format ist Abhängig von der jeweiligen Implementierung, eine bereits genannte Möglichkeit wäre JWT. Der Access Token ist zeitlich begrenzt gültig.

Refresh Token: Ein Token welches langlebig ist, also eine lange Gültigkeit besitzt. Dieses kann allerdings im Gegensatz zum Access Token widerrufen werden. Ebenfalls wird es verwendet, um ein neues Access Token vom Authorization Server anzufordern. Dafür ist keine Übergabe der Benutzerdaten nötig.

Resource Server: Eine Resource auf die nur zugegriffen werden kann, wenn ein valides Access Token vorliegt, dies könnte z.B. ein Micoservice sein.

Client: Ein Client möchte Zugriff auf den Resource Server. Clients können beispielsweise Drittanwendungen, Webanwendungen oder mobile Applikationen sein.

Client: Ein Client möchte Zugriff auf den Resource Server. Clients können beispielsweise Drittanwendungen, Webanwendungen oder mobile Applikationen sein. Sie werden auch Resource Owner genannt.

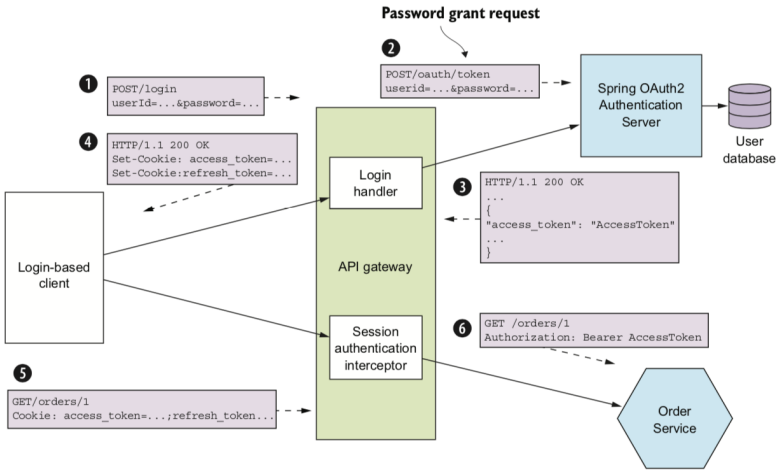

Abbildung 9: Ablauf eines sogenannten ‘password grant‘[8]

Die Abbildung 9 zeigt einen Ablauf, bei dem ein Client sein Usernamen und Passwort übermittelt (1). Diese Anfrage wird von dem API-Gateway weitergeleitet an einen Authorization Server (2). Wurde der Nutzer gefunden, wird ein Access und Refresh Token über das API-Gateway (3) an den Nutzer übermittelt (4). Mit der entsprechenden Authorisierung ruft der Client seine Bestellungen (5) ab. Dieser Request wird entsprechend von dem API-Gateway weitergeleitet (6).

Neben dem *password grant Flow* existiert auch der sogenannte *client credentials grant Flow*. Dieser funktioniert ähnlich, nur dass nicht Username und Passwort übertragen werden, sondern der Anwendungen durch gemeinsam bekannte Daten zwischen Anwendungen und Authorization Server vertraut wird. Den *client credentials grant Flow* sollte man verwenden, wenn eine Applikation Ressourcen aufrufen möchte, die außerhalb eines User-Kontextes liegen, d.h. es handelt sich in der Regel um recht allgemeine Daten.[7][8]

Aufgrund dieser Grundlage wird im Konzept VERLINKUNG-EINFÜGEN herausgearbeitet, welche Möglichkeiten zur technischen Umsetzung dieser Flows bereitstehen.

3.7 Docker

Über Docker können Anwendungen innerhalb eines Containers laufen. Ein Container ist vergleichbar mit einer sehr leichtgewichtigen, modularen virtuellen Maschine.[12] Docker in der Tiefe aufzuarbeiten, würde den Umfang dieser Arbeit überschreiten. Dennoch wird im Folgenden erläutert, warum speziell Docker sich sehr gut für Microservices eignet und wie das grundsätzliche Wirkungsprinzip von Docker funktioniert.

Anfangs wurden Microservices so definiert, dass sie als möglichst eigenständige, deploybare Einheiten zu betrachten sind. Beim Hosten - also dem Bereitstellen des Services - sollte dies ebenfalls berücksichtigt werden. Nimmt man an, man hostet alle Services

auf einer Maschine, läuft man Gefahr, dass die Microservices sich z. B. durch Portkonfigurationen oder dem Zugriff auf selbe Ressourcen behindern.[14]

Bisherige Lösungen - und je nach Anwendungsfall auch sinnvoll - bieten sich an dieser Stelle virtuelle Maschinen (**VM**) an. D.h. auf einem Host könnten mehrere Betriebssysteme laufen, die sich die Hardwareressourcen vom Host teilen. Die angestrebte Isolation zwischen den Microservices wäre erreicht und individuelle Konfigurationsmöglichkeiten wie z.B. Portfreigeben könnten sich gegensetitig nicht mehr stören. Durch die VMs entstehen allerdings eine Performanceeinbußen, welche nicht im Verhältnis zum Nutzen stehen und zusätzlich ein höherer Verbrauch der Hardwareressourcen.[14] Gesucht ist dementsprechend eine Technologie, die es schafft Services zu isolieren und dabei gleichzeitig leichtgewichtig zu sein: Docker.

Startet man einen Microservice über Docker ist dies damit gleichzusetzen, als würde im Betriebssystem der Service als Prozess gestartet werden. Es entsteht kein deutlich sichtbarer Overhead in Bezug auf Ressourcenverbrauch.[14]

Um den Docker-Aufbau (siehe Abbildung 10) besser zu verstehen, ist folgendes Vokabular nützlich:[14]

Docker-Image: Aus einer Anwendung kann man ein Image erzeugen, so dass es im Docker-Container gestartet werden kann.

Docker-Container: Wenn ein Image ausgeführt wird, läuft es in einem Container, welcher verschiedene Eigenschaften hat, wie z.B. ein eigenes Netzwerk-Interface und ein Dateisystem.

Dockerfile: Ein Dockerfile ist ähnlich einem Bauplan. Er beschreibt, wie das Image gebaut werden muss, damit es in einem Docker-Container laufen kann.

Repository: Ein Repository speichert Images. In der Regel hält ein Repository mehrere Images von derselben Anwendung mit verschiedenen Versionen bereit.[12]

Docker-Registry: Eine Docker-Registry verwaltet mehrere Repositories.

Docker-Host: Ein Docker-Host unterstützt die Docker-Technologie, d.h. es können entsprechend Images ausgeführt werden. Zusätzlich ist es möglich, Images von einem Repository auszuführen. D.h. man muss die Images nicht extra auf den Host laden, um sie auszuführen.

Bei genauerer Betrachtung der Abbildung 10 fällt auf, dass Container in einem Netzwerk sind und sich denselben Kernel teilen. Dateisysteme, Netzwerk-Interface und containereigene Prozesse sind voneinander isoliert. D.h. Portfreigaben oder ähnliches behindern sich nicht.

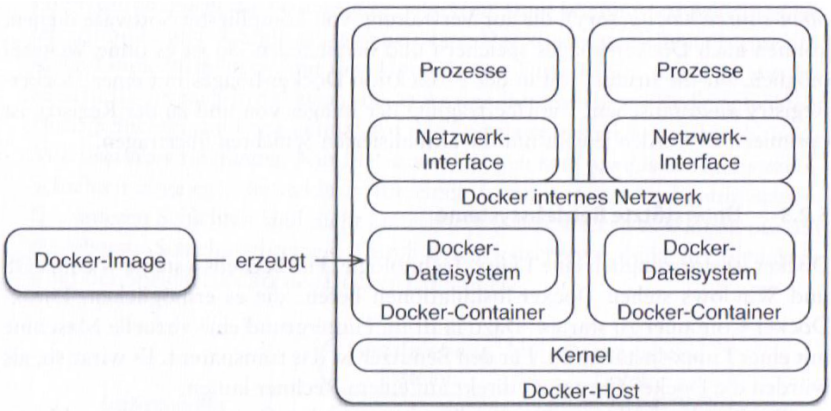

Abbildung 10: Docker Architektur[14]

Zusammenfassend lässt sich sagen, dass durch das Nutzen von Docker eine isolierte und ressourcenvertretbare Trennung von Microservices auf einem Host erreichen lässt.

4 Konzept

Das ist der Konzeptpart

4.1 Anforderungen definieren

4.2 Macroarchitektonische Festlegungen

- MySQL, Docker, selbe Begründung - REST (Verlinkung auf Mobile Development-Modul) - Bounded Context erstellen - Aus sicht des User Story Mappings:

4.3 Wahl des API-Gateway

- Ocelot - welches gibt es noch? - Eigenbau - Auth0

4.4 Wahl der Kommunikation

- Erwähnen, dass asynchrone Kommunikation verwendet wird, aufgrund der Vorteile aus den Grundlagen - Technologie: Masstransit (RabbitMQ) vs. Kafka

4.5 Wahl der Authenfizierung/Authorisierung

- OAuth2: Authentication Framework (mit OpenId Connect?) -> Spricht die Zeit, Komplexität und so eine Komplexität gar nicht benötigt - JWT: Also Protocol - Dritteranbieter Dienst -> Auth0 (zu teuer)

5 Implementierung

Das ist der Implementierungspart

5.1 Service Architektur

// Grafik erstellen, welche Services es gibt und wie diese Kommunizieren // Docker erwähnen und Beispiel einfügen

5.2 Umsetzung API-Gateway

// Grafik erstellen, welche Services es gibt und wie diese Kommunizieren

5.3 Authentifizierung und Authorisierung

// Authentifizerungsart umsetzen

5.4 Asynchrone Kommunikation

// ieine Queue umsetzen (RabbitMQ, Kafka...)

6 Fazit

Gelungen, weil

6.1 Ausblick

- Loadbalancing - ELK Stack - Service Mash - Kubernetes (Orchestrierung) - Sichtbarkeit von Microservices (was soll alles erreichbar sein): Also nur das Gateway und/oder auch die Microservices

Tabellenverzeichnis

Abbildungsverzeichnis

Literatur

[1] Mark Fussel (msfussell). *Einführung in Microservices in Azure - Microsoft-Dokumentation*. abgerufen am 9.4.2019. 2017. URL: https://docs.microsoft.com/de-de/azure/service-fabric/service-fabric-overview-microservices.

[2] Ferdinand Birk. „Microservices - Eine State-of-the-Art Bestandsaufnahme und Abgrenzung zu SOA“. Universität Ulm, 2016.

[3] Sebastian Gauder. „A competitive food retail architecture with microservices“. 2019.

[4] Matt McLarty Irakli Nadareishvili Ronnie Mitra und Mike Amundsen. *Microservice Architecture - Aligning Principels, Practices, and Culture*. O Reilly, 2016.

[5] Paulo A. Pereira Morgan Bruce. *Microservices in Action*. Manning, 2019.

[6] Dennis Peuser Nhiem Lu Gert Glatz. *Moving mountains – practical approaches for moving monolithic applications to Microservices*. University of Applied Science und Arts, Dortmund, Germany, adesso AG, Dortmund, Germany. 2017.

[7] Aaron Parecki. *OAuth 2.0 Client Credentials Grant*. abgerufen am 11.5.2019. o. J. URL: https://oauth.net/2/grant-types/client-credentials/.

[8] Chris Richardson. *Microservice Pattern*. Manning, 2019.

[9] Golo Roden. *Domain-driven Design erklärt*. abgerufen am 15.3.2019. 2016. URL: https://www.heise.de/developer/artikel/Domain-driven-Design-erklaert-3130720.html?seite=all.

[10] Cesar de la Torre (cesardelatorre). *Implementieren ereignisbasierter Kommunikation zwischen Microservices*. abgerufen am 16.4.2019. 2018. URL: https://docs.microsoft.com/de-de/dotnet/standard/microservices-architecture/multi-container-microservice-net-applications/integration-event-based-microservice-communications.

[11] o. V. *Plan an API Gateway system.* abgerufen am 1.5.2019. o. J. URL: https://docs.oracle.com/cd/E55956_01/doc.11123/administrator_guide/content/admin_planning.html.

[12] o. V. *Was ist Docker?* abgerufen am 11.5.2019. o. J. URL: https://www.redhat.com/de/topics/containers/what-is-docker.

[13] Eberhard Wolff. *Das Microservice Praxisbuch - Grundlagen, Konzepte und Rezepte.* dpunkt.verlag, 2018.

[14] Eberhard Wolff. *Flexible Software Architectures.* Leanpub, 2016.