



UNIVERSITATEA DIN
BUCUREȘTI

FACULTATEA DE
MATEMATICĂ ȘI
INFORMATICĂ



SPECIALIZAREA TEHNOLOGIA INFORMAȚIEI

Lucrare de licență

DEVOPS ÎN MICROSERVICII

Absolvent

Bahrim Dragoș

Coordonator științific

Conferențiar Doctor Kevorchian Cristian

București, iulie 2023

Rezumat

Microserviciile sunt o arhitectură de sisteme distribuite ce a luat amploare în urma structurării organizațiilor către piață. Aceasta a fost facilitată de schimbări în modul în care lansăm produsele prin dezvoltarea platformelor cloud și a ce a dus la îmbunătățiri la timpul de livrare dar și a consistenței. Obiectivul este prezentarea acestor concepte și implementarea acestora prezentând modul de gândire ce influențează ciclul produsului. În acest scop, mă documentez legat de domeniu și încerc să îmi fac o idee de ansamblu asupra aspectelor ce trebuie luate în construirea unui sistem iar apoi o să aplic aceste informații. Aceasta lucrare ar trebui să servească ca un prim contact cu microserviciile dar și a practicilor DevOps.

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce vitae eros sit amet sem ornare varius. Duis eget felis eget risus posuere luctus. Integer odio metus, eleifend at nunc vitae, rutrum fermentum leo. Quisque rutrum vitae risus nec porta. Nunc eu orci euismod, ornare risus at, accumsan augue. Ut tincidunt pharetra convallis. Maecenas ut pretium ex. Morbi tellus dui, viverra quis augue at, tincidunt hendrerit orci. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam quis sollicitudin nunc. Sed sollicitudin purus dapibus mi fringilla, nec tincidunt nunc eleifend. Nam ut molestie erat. Integer eros dolor, viverra quis massa at, auctor.

Cuprins

1	Introducere	5
2	Preliminarii	8
3	Microservicii	9
3.1	Concepte generale	9
3.2	Design	12
3.2.1	Interconectare	12
3.2.2	Domain Driven Design	13
3.2.3	Concepte de programare orientată obiect în microservicii	14
3.2.4	Migrarea	15
3.3	Tehnici de implementare	17
3.3.1	Tipuri de comunicare	18
3.3.2	Tehnologii de comunicare	20
3.3.3	Schimbări în funcționalitate	22
3.3.4	Partajare între microservicii	23
3.4	Build	27
3.5	Testare	27
3.6	Deployment	27
3.7	Monitorizare	27
3.8	Securizare	27
3.9	Evoluție	27
4	Orchestrare	28
4.1	Istoric deployment	28
4.2	Tipuri de deployment	28
4.2.1	Baremetal	28
4.2.2	Mașină virtuală	28
4.2.3	Container	28
4.2.4	Function as a Service	28
4.2.5	Platform as a Service	28

4.2.6	Container as a Service	28
4.3	Docker	28
4.4	Docker Swarm	28
4.5	Kubernetes	28
4.6	Red Hat OpenShift	28
4.7	HashiCorp Nomad	28
4.8	Infrastructure as a Service	28
5	Îmbunătățirea metodelor de dezvoltare software	29
5.1	Schimbari in livrarea produselor	29
5.2	DevOps și aspecte ale comunității	29
5.3	Ce înseamnă să aplici DevOps	29
5.4	Unelte	29
6	Aplicație	30
7	Concluzii	31
	Bibliografie	32

Capitolul 1

Introducere

Domeniul informaticii, calculatoarelor și al tehnologiei informației este unul care va fi mereu în continuă evoluție. Dorința companiilor de a-și servi clienții cât mai rapid a dus la un avans tehnologic în care mereu apare ceva nou ce are ca scop îmbunătățirea proceselor actuale. Această arie și-a început dezvoltarea aproape acum o sută de ani iar avansul poate fi apreciat numai gândindu-ne la modul în care scriam cod și livram produse acum, acum zece ani și acum două zeci de ani. Se observă o diferență și la accesibilitatea unităților de calcul, actual majoritatea persoanelor au acces la un dispozitiv ce se poate conecta la Internet, ceea ce le permite să devină consumatori la diferite servicii.

Piața pentru unități de calcul a evoluat. În 1943, Thomas Watson, director la IBM menționează că „I think there is a world market for maybe five computers” (Cred că există o piață globală pentru poate cinci computere). În 1999, Michael Barr afirmă în „Programming Embedded Systems in C and C++” „One of the more surprising developments of the last few decades has been the ascendance of computers to a position of prevalence in human affairs. Today there are more computers in our homes and offices than there are people who live and work in them. Yet many of these computers are not recognized as such by their users.” (Una dintre cele mai surprinzătoare dezvoltări a ultimelor decenii a fost ascensiunea computerelor într-o poziție predominantă în afacerile oamenilor. Astăzi sunt mai multe computere în casele noastre și în birouri decât persoane care trăiesc și lucrează în ele. Însă multe dintre acestea nu sunt recunoscute de către utilizatorii lor). Cele două fraze evidențiază modul cum s-a schimbat prezența computerelor în viața noastră. La început acestea erau foarte costisitoare, greu de administrat și de operat însă acestea au devenit din ce în ce mai mici iar recent aproape orice are nevoie de semiconductori datorită integrării ce le fac „smart”.

Aceste afirmații se aplică și pentru produsele software, ce au ca menire să ofere asistență altor aplicații sau să fie consumate direct. Indiferent de locul în care mă duc, probabil persoana cu care aş interacționa folosește un computer, fie că îmi cumpăr ceva de la un magazin făcând o plată cu cardul sau că mi se livrează un colet iar curierul marchează pe AWB că a fost livrat. Christopher Little afirmă că „Every company is a technology com-

pany, regardless of what business they think they are in. A bank is just an IT company with a banking license.” (Orice companie este o companie ce se axează pe tehnologie, indiferent de mediul de afaceri în care se află. O bancă este doar o companie IT cu licență de a funcționa ca o bancă.)

Desigur nu toate companiile sunt la fel și nu toate concurează cu toate companiile, însă chiar dacă nu ești o companie ce activează în tehnologie, probabil folosești componente tehnologice pentru a-ți îmbunătăți randamentul sau să oferi clienților un avantaj față de competiție, care probabil gândește în același mod.

Însă pentru fiecare companie obiectivul este identic, să livreze clienților produse de calitate și cât mai rapid. Pentru aceasta trebuie să ne asigurăm că avem inginerii necesari pentru implementarea cererilor noi, însă momentul în care aceasta este terminată este doar începutul procesului de integrare și ulterior lansare, și uneori acesta este cel care reprezintă un blocaj din mai multe puncte de vedere ce limitează numărul de câte ori putem aduce o îmbunătățire produsului nostru.

În încercarea de a livra mai rapid sau pentru a ușura modul în care manevrăm un influx sau o lipsă de trafic, s-a încercat folosirea unei arhitecturi paralele asupra aplicațiilor creând arhitecturi orientate pe servicii, iar în momentul în care aceste servicii sunt concentrate pe un număr redus de funcționalități și pot fi lansate independent și să își păstreze funcționalitate putem vorbi de o arhitectură bazată pe microservicii.

Însă schimbând arhitectura sau practiciile din modul de desfășurare al livrării produsului nu este suficient pentru a asigura performanța, întrucât ambele dintre ele aduc dezavantaje ce trebuie tratate iar uneori acestea sunt mai mari decât avantajele pe care schimbarea le-ar aduce, însă implementarea corectă în locurile ce ar beneficia de o astfel de abordare poate să aducă performanțe ce nu ar fi posibile cu procesele vechi.

Afinitatea mea pentru microservicii provine de la lipsa de cunoștințelor pentru a putea alege dintre mai multe limbaje de programare, framework-uri sau platforme. De exemplu, ce avantaje mi-ar face să aleg ca pentru server-ul meu să folosesc o aplicație în Node.JS sau una în Go? Chiar dacă putem să cunoaștem niște lucruri informative la început, cei mai buni indicatori sunt monitorizările proprii în producție. De asemenea pot fi cazuri în care îmbunătățirile aduse unei platforme pe parcurs o fac mai bună față de ce era inițial (dezvoltarea TypeScript poate să influențeze alegerea inițială a unei platforme SpringBoot ce este type-safe). Astfel, pe o arhitectură bazată pe microservicii putem să folosim limbajul potrivit pentru serviciul căruia îi aduce cele mai multe beneficii. Acest lucru se extinde și pentru baze de date. Poate în timpul evoluției serviciului apar tehnologii noi, însă având un singur lucru central ar face migrarea mult mai grea, de asemenea uneori ar trebui să facem compromisuri pe care alegerea unui alt tip de baze de date ar face să o dispară.

Dezvoltarea tematicilor DevOps pornește de la ușurința cu care ne putem crea o bază de date folosind containere, întrucât instalarea unei baze de date pe calculatorul

personal în timp ce lucram la proiectele de facultate a făcut să am experiențe în care la finalul fiecărui an îmi reinstalam sistemul de operare doar pentru că o dată ce instalam o bază de date crea suficiente servicii, foldere care după o dezinstalare încă rămâneau. Folosind Docker, pot crea o bază de date cu o singură linie de cod, care în același timp să fie preconfigurată și să o pot împărtăși cu colegii de proiect ca să nu pierdem timp în configurarea mediului de lucru.

Lucrarea mea va urma o structură în care prezint noțiunile teoretice la început iar la final încerc să aplic aspectele prezentate în crearea unei aplicații. Inițial pornesc cu prezentarea microserviciilor continuând cu diferitele moduri în care acestea pot fi lansate, imediat după cu dezvoltarea noțiunilor de DevOps ce au ca scop accelerarea dezvoltării, iar la final o prezentare generală a modului în care îmi construiesc aplicația.

Capitolul 2

Preliminarii

Cel puțin până la momentul la care am decis ce temă să aleg pentru lucrarea mea de licență niciun curs de bază nu a abordat în detaliu tema pe care vreau să o dezvolt, din acest motiv lucrarea mea de licență nu este doar aprofundarea unui concept și crearea unei aplicații, ci încercarea familiarizării cu obiectele de lucru în același timp. Desigur, unele aspecte s-ar putea desprinde și în urma practicii în industrie dar nu a fost suficient ca să pot să consider că ma cunosc de bază în acest domeniu și doar încerc să găsesc lucruri cât mai specifice.

De asemenea, microserviciile sunt o arhitectură destul de nouă ce a luat amploare datorită limitărilor din punctul de vedere al dezvoltării și scalării aplicațiilor monolitice, astfel în opinia mea, este destul greu să vizualizezi anumite lucruri fără experiență la prima mână cu limitările acestea. Întrucât serviciul pe care îl dezvolt nu ar necesita neapărat o arhitectură bazată pe microservicii, ar funcționa complet normal și aș reduce foarte mult din complexitate dacă ar fi dezvoltată în mod normal.

La fel și cu practiciile DevOps, acestea se axează mai mult companiilor care au nevoie de timp de penetrare a pieții foarte rapid, însă eu doar încerc să mă familiarizez cu setarea și folosirea serviciilor pentru uzurile mele proprii. Un lucru important ar fi ca ce folosesc să nu mă limiteze, de exemplu infrastructura de care am nevoie va proveni din Azure doar pentru că primesc credite gratuite ca și student, însă în absența lor probabil aș încerca lucruri cu cost minimal sau chiar nici să nu le fac, și probabil aș avea același randament.

Capitolul 3

Microservicii

3.1 Concepte generale

Microserviciile sunt o arhitectură de sisteme distribuite ce a devenit populară în ultimele decenii datorită necesității scalării. Ceea ce o separă de o arhitectură orientată pe servicii este faptul că microserviciile sunt mult mai specifice, acoperind o singură parte din afacere și faptul că pot fi lansate în execuție independent, astfel microserviciile ar trebui să depindă cât mai puțin de altele. Comunicarea se face direct către interfețele expuse de fiecare, iar informațiile reținute de fiecare (de exemplu baza de date de date sau diferite fișiere) pot să fie modificate doar prin metodele de comunicare oferite ci nu direct.

Un concept important în microservicii este ascunderea de informații, astfel din exterior fiecare ar trebui să fie tratat ca o cutie neagră ce face anumite lucruri și expune anumite date, nu contează detaliile de implementare și nici tehnologiile folosite în implementare, ci doar capabilitățile fiecăruia. Acest lucru permite să modificăm microserviciul în funcție de cerințe, chiar și refactorizarea metodelor inițiale cât timp se respectă capabilitățile pe care microserviciul ar trebui să le îndeplinească.

Cel mai important aspect al microserviciilor este capacitatea de a lucra independent de celelalte. Acestea trebuie să funcționeze independent de celelalte iar dacă aducem schimbări într-unul atunci nu suntem obligați să facem schimbări în alt microserviciu, în caz contrar ar apărea dificultăți în lansare. Obținerea independenței poate să aducă contribuții majore în timpul de deployment însă implementarea este mult mai complicată datorită necesității comunicării cu alte servicii.

Un alt aspect de bază este acoperirea unui singur segment din afacere prin intermediul unui microserviciu. Alături de caracteristica precedentă, dacă afacerea și-ar dezvolta necesitățile ar fi mult mai dificil să modificăm mai multe microservicii și ulterior să coordonăm lansarea lor.

Independența microserviciilor este importantă, în acest scop, ele ar trebui să fie singu-

rele care ar putea să își modifice starea. Întrucât starea se referă la elementele componente, de exemplu o bază de date, acestea nu ar trebui să fie împărtășite și nici accesate de alte microservicii. De asemenea ar trebui evitate modificarea interfețelor declarate pentru a evita necesitatea modificării altor microservicii. „Dimensiunea microserviciilor ar trebui să fie redusă prin oferirea unui număr restrâns de funcționalități declarate pentru a fi ușor de înțeles și întreținut. Însă important este cât de multe microservicii pot fi administrate întrucât un număr mai mare de microservicii, deși mici ca funcționalități, aduc dificultăți în comunicare, scalare și depanare.

Microserviciile sunt agnostice din punct de vedere al tehnologiilor care ar putea fi folosite. Din acest motiv, acestea oferă flexibilitate însă prețul plătit este crearea a mai multor puncte vulnerabile.

Orice prezentare al acestui tip de arhitectură nu ar fi completă fără prezentarea conceptului de „monolith” (monolit), întrucât este considerată o arhitectură veche reprezentând modul de funcționare din trecut. În cel mai simplist mod de funcționare, sistemul monolitic este reprezentat de un singur proces „gigantic” ce acoperă toată funcționalitatea sistemului, în general conectat la o singură bază de date ce și ea este reprezentată de un singur proces. Putem intui problemele cu această organizare, dacă conexiunea către procesul monolitic cade sau baza de date are probleme, întreg sistemul este afectat. Microserviciile încearcă să acopere aceste probleme prin independență, astfel dacă un microserviciu este căzut, celelalte nu sunt afectate. Procesul poate fi însă separat pe module dezvoltate separat însă care la final se cuplează formând un singur proces. Există conceptul de sistem monolitic distribuit în care acesta este la fel separat în module independente însă care nu ar funcționa dacă nu sunt toate active, acest tip de sistem oferă toate dezavantajele sistemelor distribuite dar și a monolitului ceea ce îl face destul de rar întâlnit.

Chiar dacă o arhitectură monolitică prezintă dezavantaje evidente ce sunt ușor de văzut, acestea nu sunt la fel de mari pentru companiile mici. Numărul redus de aplicații ce trebuie lansate și monitorizate, posibilitatea de reutilizare a codului mult mai ușoară, obținerea infrastructurii pentru proces și uneori chiar și scalarea, deși este limitată poate să facă o arhitectură monolitică mult mai atractivă. Nu ar trebui să asociem arhitectura monolitică ca un lucru antic, ci să o considerăm ca o opțiune. Însă pentru o organizație mică, acesta ar trebui să fie doar un punct de plecare, eventual după ce serviciul oferit devine mai popular s-ar putea să ajungem să cunoaștem limitările arhitecturii și să ne orientăm tot către microservicii, însă dacă nu avem succes, eliminăm multe complexități apărute în urma introducerii microserviciilor.

Principalele avantaje ale microserviciilor provin de la baza acestora, o arhitectură distribuită, însă cuplat cu conceptul de ascundere a informației și domain-driven design (design orientat domeniu) aduc multe alte avantaje asupra altor arhitecturi distribuite. În cadrul unui sistem format din microservicii putem folosi orice tip de tehnologie pentru

program și orice tip de bază de date, întrucât ascundem implementarea de exterior. Astfel putem alege tehnologii ce aduc avantaje în dezvoltarea microserviciului. Acest avantaj duce la capacitatea de a folosi tehnologii noi fără a afecta tot sistemul, însă limitează capacitatea de a împărtăși cod (de exemplu prin librării interne, întrucât acestea ar trebui să fie rescrise pentru fiecare limbaj). Un sistem distribuit rezistă mult mai ușor la căderi, întrucât acestea pot fi tratate ca să se prevină un lanț, însă cu un număr mare de puncte slabe devine greu de aproximat cât de mult este afectat un sistem. Scalarea unui sistem monolitic este ușoară, doar replicăm procesul, însă la microservicii scalarea poate fi mult mai concentrată pe punctele care chiar au nevoie să fie scalate. Lansarea unei schimbări într-un microserviciu este mult mai ușoară întrucât nu necesită crearea de timp mort în aplicație, întrucât nu tot sistemul este înlocuit ci doar o mică parte din acesta, ce poate fi chiar și mai controlată prin diferite aplicații de orchestrare.

Adaptarea microserviciilor vine cu un cost, pe lângă dificultățile de implementare a design-ului apar și alte probleme, de exemplu dezvoltarea locală. Atunci când dezvoltăm un microserviciu acesta, posibil, este nevoit să comunice cu alte microservicii care la rândul lor comunică cu alte microservicii și putem continua, toate acestea microservicii trebuie să funcționeze în același timp pe laptop-ul dezvoltatorului, întrucât nu putem folosi microserviciile din producție, însă nu e posibil să rulăm foarte multe microservicii pe un singur computer depinzând de configurație. Adoptarea microserviciilor, în general nu poate avea succes decât dacă este combinată cu elemente de DevOps, ceea ce înseamnă ca dezvoltatorii și operatorii să învețe tehnologii noi, ceea ce este un impediment în momentul în care vrem să lansăm un produs rapid în piață și nu avem experiența necesară. Microserviciile necesită mult mai multă infrastructură pentru a fi rulată, mai multă tehnologie ce trebuie să fie folosită pentru ca dezvoltarea și livrarea să decurgă eficient, însă pe termen lung acesta poate să fie prevăzută dacă arhitectura este folosită cum trebuie prin livrare mai rapidă și mai eficientă deci profitul ar putea să crească. Monitorizarea sistemului devine mai complicată, într-un proces monolitic toate fișierele de jurnalizare (logging) sunt în același loc. Într-un sistem distribuit, acestea trebuie să fie colectate, și ulterior ansamblate pentru a avea sens, acest lucru devine mai dificil când intră în discuție fenomenul de replicare al unei instanțe. Securizarea unui sistem distribuit este mai grea, traficul de date se face mult mai mult prin rețele sau prin diferite căi alternative ceea ce necesită un grad ridicat de atenție. Testarea unui sistem format din microservicii este mult mai dificilă, mai ales când trebuie testate mai multe microservicii în același timp, acesta duce la creșterea timpului necesar pentru a primi răspuns. Sistemele distribuite cresc timpul de răspuns întrucât acestea nu se mai face în cadrul unui singur proces, deci datele trebuie codificate și decodificate atunci când trec dintr-un mediu în altul. În cadrul unui sistem distribuit o altă problemă care apare este consistența datelor, dată de faptul că în timpul unei cereri, datele dintr-un microserviciu se pot schimba, astfel tranzacțiile devin dificile mai ales când înainte ne bazam pe atomicitatea și consistența oferită de

baza de date.

Astfel, microserviciile funcționează cel mai bine pentru organizațiile mari ce vor ca dezvoltatorii lor să livreze facilități noi foarte repede care să nu fie întârziate de compatibilitatea cu caracteristici deja existente. Prin natura variată a acestei arhitecturi se pot folosi tehnologii noi dar și folosirea platformelor cloud, astfel se oferă multă flexibilitate cu prețul creșterii complexității. Pentru companiile mici, al căror rată de succes este greu de calculat, overhead-ul produs de adaptarea unei arhitecturi în acest stil nu ar aduce suficient de multe avantaje și ar trebui să opteze pentru o arhitectură tradițională ce este stabilă și oferă colaborare mult mai ușoară în cadrul aceluiași proiect la scale mici.

3.2 Design

Partea de creare a schemei arhitecturale a unui sistem este una dintre cele mai importante faze ale dezvoltării. Acesta este momentul în care putem descoperii că microserviciile nu ar fi chiar o alegere potrivită pentru cerințele noastre. Din acest motiv este important să nu ignorăm potențiale probleme ci să se formuleze toate problemele pe care le putem aștepta în implementare

3.2.1 Interconectare

Deși microserviciile pot fi lansate și pot funcționa independent, acestea trebuie să funcționeze în cadrul unui sistem complex. Astfel trebuie definite modul de comunicare și interfețele expuse. În definirea interconectării întâlnim câteva concepte de bază prezentate anterior ce provin din descompunerea în module.

Ascunderea de informații este procesul de separare al detaliilor de implementare pentru a evita modificarea în alte locuri în cazul în care acestea se schimbă. Acest concept oferă varii beneficii precum îmbunătățirea timpului de dezvoltare datorită posibilității de dezvoltare în paralel, creșterea înțelegerii sistemului mult mai ușor întrucât fiecare modul poate să fie înțeles independent și flexibilitate oferit de faptul că putem schimba funcționalitate fără a fi necesar ca alte părți ale sistemului să fie modificate. David Parnas spune că „The connections between modules are the assumptions which the modules make about each other.” (Conexiunile create între module sunt ipotezele create de module față de celelalte) [4]. Astfel, cu cât alte microservicii folosesc un anumit microserviciu mai mult se crează mai multe locuri în care devine mai dificil de modificat, astfel cu cât întâmpinăm problema aceasta mai puțin cu atât putem lansa acel microserviciu fără a lua în considerare modificări care să pastreze compatibilitatea cu elementele existente.

Un alt concept important este coeziunea, scopul microserviciilor este de a funcționa independent, astfel vrem ca lucrurile care interacționează să fie grupate iar atunci când se schimbă ceva să nu trebuiască să modificăm multe locuri. Elementele care nu

interacționează ar trebui să fie separate pentru a nu crea probleme de regresie.

În sisteme distribuite apare cuplarea, atunci când un microserviciu depinde mai mult sau mai puțin de altul pentru a avea funcționalitate completă. Un sistem cu cuplare mică poate să fie lansat fără să trebuiască modificări în locuri multiple.

Cele două concepte definesc lucruri asemănătoare, astfel ele se pot rezuma ușor ca „A structure is stable if cohesion is strong and coupling is low.” [2] (O structură este stabilă dacă coeziunea este ridicată și cuplarea este scăzută). Coeziunea se referă la lucrurile din interior iar cuplarea se referă la exterior iar acest lucru este greu de balansat.

Cuplarea poate apărea sub diferite forme, cuplare de domeniu în care un microserviciu interacționează cu alt microserviciu pentru că are nevoie de funcționalitate ce depășește atribuțiile lui. Astfel de cuplare, în general, nu poate fi evitată într-un sistem distribuit. Acesta este cel mai scăzut nivel de cuplare însă trebuie să evităm cazul în care un microserviciu depinde un centru pentru alte microservicii întrucât ar însemna ca acesta are prea multe responsabilități. Există conceptul de cuplare temporară, adică atunci când un microserviciu așteaptă ca alt microserviciu să execute ceva, astfel cuplarea se întâmplă pentru că cele două operațiuni se întâmplă în același timp. Cuplarea Pass-Through apare atunci când un microserviciu este nevoit să trimită date unui alt microserviciu întrucât acesta trebuie să trimită datele unui microserviciu la un nivel mai mic, astfel un microserviciu are rolul doar de a transmite date. Acest lucru este problematic întrucât dacă formatul acestor date ar trebui să se schimbe, ar trebui să modificăm mai multe microservicii. Evitarea acestui tip de cuplare se face prin schimbarea modului prin care comunicăm de exemplu microserviciul ce emite date ar putea să trimită direct datele către microserviciul ce are nevoie de ele sau ca microserviciul intermediar să trimită datele în format pur, fără să îl intereseze o formă anume. Cuplarea comună apare atunci când mai multe microservicii încearcă să acceseze o resursă comună, de exemplu o bază de date sau locația unui fișier. Problema în acest caz este că dacă modul în care această resursă interacționează se schimbă, mai multe microservicii trebuie modificate. Sau dacă unul din microservicii are nevoie de un comportament special, alte microservicii ar trebui să se adapteze ceea ce crează probleme în dezvoltare. Soluții posibile de evitare pot fi crearea unui alt microserviciu ce funcționează ca o interfață pentru resursa comună ce va funcționa după principiile unui microserviciu. Cuplarea de conținut apare atunci când un microserviciu modifică starea internă a altui microserviciu, de exemplu prin modificarea bazei de date a acesteia. Acest lucru este periculos întrucât pot apărea modificări nedorite prin utilizare incorectă, astfel se încalcă procesul de ascundere de informații.

3.2.2 Domain Driven Design

Domain Driven Design este un concept introdus de Eric Evans în cartea sa [3]. Acesta se bazează pe crearea codului în jurul afacerilor pentru ca acesta să fie mai ușor de

înțeles prin creșterea interacțiunii cu analiștii de afaceri, acest lucru este benefic pentru microservicii întrucât acestea se pot orienta către mediul de afaceri pentru a se dezvolta alături de acesta fiind mult mai eficient.

Unul din concepte este limbajul universal ce cuprinde ca codul din interiorul unui microserviciu să folosească termeni folosiți de către utilizatori și de către analiști, acest lucru este benefic întrucât face comunicarea mult mai eficientă, o persoană nou venită poate să înțeleagă afacerea și codul în același timp. Dacă cele două sunt diferite, atunci pot exista probleme în comunicarea problemelor.

În DDD, un agregat este o unitate ce are o anumită stare ce se poate modifica în timp și are ca scop modelarea unui concept din domeniul real. În general un agregat este cuprins în întregime de către un microserviciu chiar dacă acesta poate conține mai multe microservicii. Un agregat poate fi schimbat din exterior însă poate să își controleze singur starea pentru a preveni stările ilegale

Mărginirea contextului se referă la ascunderea mai multor activități ce țin de implementare cu scopul de a defini anumite responsabilități sigure pe care le îndeplinește. Acest lucru se poate îndeplini prin ascunderea modelului, atunci când vrem să expunem o entitate nu trebuie să oferim toate informațiile despre acesta ci doar cele necesare. Distribuirea unui model, adică a unei entități între mai multe microservicii se face prin traducerea acestei entități în domeniu propriu.

DDD în microservicii este un concept fundamental întrucât ambele pun la centru afacerea și dezvoltarea în jurul ei. Astfel conceptele prezentate în DDD pot fi aplicate microserviciilor mărginirea contextului are ca bază ascunderea de informații și duce la crearea de formate standard.

3.2.3 Concepte de programare orientată obiect în microservicii

Multe din problemele cu care ne confruntăm în microservicii pot fi tratate folosind concepte ce au ca scop ușurarea modului în care creăm sisteme prin intermediul claselor în programarea orientată obiect. Deși diferite în natură, ambele au ca rol definirea unor interacțiuni între elemente ce sunt făcute să funcționeze separat sau prin intermediul unuor interfețe.

Începând cu conceptele de bază, abstractizarea se referă la ascunderea implementărilor interne și afișarea doar facilitățile oferite de clasă ce pot fi accesate din exterior. Acest lucru face referire la conceptul de ascundere al informațiilor descris anterior.

Encapsularea se referă la agregarea de date și attribute, adică funcționalități pentru a crea un tot unitar, acest principiu este compatibil cu microserviciile și nevoia acestora de a funcționa independent folosind datele proprii și permite o securizare mai bună a stării interne.

Moștenirea se referă la capacitatea unei clase de a prelua attributele și funcționalitățile

unei clase deja existente, deși mult mai greu de imaginat în microservicii, moștenirea în microservicii ne permite să creăm microservicii ce extind pe cel existent fără a cauza probleme.

Polimorfismul este un concept ce face referire la capacitatea accesării a unor multiple entități prin intermediul unei singure interfețe, în contextul microserviciilor putem spune că un microserviciu ar putea să înlocuiască alt microserviciu dacă are aceeași interfață de comunicare și duce la crearea a unor aceleași rezultate.

Pentru microservicii putem aplica și concepte mai complicate de OOP precum principiile SOLID. Principiile SOLID sunt un set de reguli cu scopul de a oferi dezvoltatorilor o metodă de a scrie cod, astfel încât acesta să respecte cerințele actuale și să permită dezvoltări/modificări ulterioare cu ușurință fără a avea dificultăți cu metodele deja existente. SOLID este un acronim iar fiecare literă face referire la un principiu. Principiul responsabilității unice (Single Responsible Principle) se referă la necesitatea ca fiecare clasă și metodă să aibă o singură funcționalitate, făcând asta reducem complexitatea dar și eventuale redundanțe create prin folosirea aceluiași cod în mai multe locuri făcând ca fiecare porțiune din cod să aibă un singur motiv de a exista. Principiul Deschis – Închis (Open Close Principle) conduce la faptul că entitățile scrise trebuie să fie deschise pentru moștenire dar să nu poată fi modificate, astfel asemănător cu principiul anterior, codul deja scris nu va fi afectat întrucât noile modificări se aduc asupra unei porțiuni noi. Principiul substituției al lui Liskov (Liskov Substitution Principle) se referă la faptul că clasele derivate trebuie să acopere funcționalitățile clasei pe care o moștenesc astfel încât acestea să poată să le înlocuiască fără modificări adiționale. Principiul separării interfețelor (Interface Segregation Principle), clasele ce implementează interfețe nu ar trebui să implementeze metode de care nu au nevoie, astfel dacă este nevoie să adăugăm o metodă nouă unei interfețe aceasta trebuie să fie implementată doar de către obiectele care au nevoie, dacă acest fapt nu este satisfăcut trebuie ca interfețele să fie separate ca să nu forțăm alte clase să implementeze metode noi. Principiul inversării dependențelor (Dependency Inversion Principle) se referă că modulele de nivel înalt nu ar trebui să depindă de cele de nivel scăzut, ele trebuie să fie abstractizate astfel acestea pot fi modificate în orice moment dacă vrem să aducem un alt modul cu funcționalități similare dar cu implementări diferite.[1]

3.2.4 Migrarea

În planificarea urbană există conceptul de dezvoltare pe spațiu verde. Aceasta se referă la dezvoltarea pe un loc în care la momentul începerii lucrărilor este spațiu verde ce nu necesită costuri adiționale pentru curățare. Atunci când vrem să construim un sistem bazat pe microservicii pornind de la zero, apare acest fenomen. Însă de multe ori există un sistem existent ce ajunge la anumite limitări ce fac dorința de trecere către

microservicii. Acest lucru face referire la conceptul de dezvoltare pe spațiu maro, adică pe un spațiu post industrial în care există costuri de demolare sau restaurare și ce necesită o atenție mult mai ridicată datorită posibilității alterării compoziției pământului ce ar putea să fie toxic.

Migrarea unei aplicații monolitice nu ar trebui să fie făcută cu scopul a obținerii unei aplicații formate din microservicii ci pentru că există o funcționalitate ce este împiedicată de a funcționa la capacitățile necesare pentru funcționare optimă, astfel migrarea unui sistem monolitic ar trebui să fie făcută incremental și posibil nu complet dacă nu este nevoie.

Începutul migrării ar trebui să fie făcut cu o analiză a sistemului și identificarea dacă avem posibilitatea de a întreține microserviciile, acest lucru înseamnă să avem prezentă infrastructură și oameni care să știe să configureze uneltele necesare precum cele de CI/CD pentru a livra cu aceeași consistență. Ulterior putem analiza sistemul pentru a observa locurile unde se formează coeziune fapt ce duce la posibilitatea creării unui microserviciu cu cod grupat. Pentru acest lucru se pot folosi unelte precum CodeScene ce pot identifica porțiunile de cod ce se schimbă împreună. Acest lucru ajută în crearea unui context mărginit și să identificăm modul în care acesta ar funcționa cu exteriorul.

O dată ce identificăm lucrurile ce vrem să le separăm putem separa funcționalități și elementele conținute, UI-ul, backend-ul și baza de date. Am fi tentați să ignorăm o parte din aceste componente însă microserviciile funcționează cel mai bine atunci când toate acestea sunt separate prin simplul motiv ca chiar dacă UI-ul nu este 1:1 cu backend-ul, acestea tot vor fi modificate în același timp. În acest scop, există două modalități în care putem face migrarea, pornind de la cod, însemnând că separăm funcționalitatea însă în continuare folosim datele din baza de date monolitică. Acest tip de abordare este destul de simplu, întrucât obținem rezultate bune pe termen scurt, însă există limitări întrucât este posibil să nu putem separa datele din baza de date, fapt ce înseamnă că munca depusă pentru separare ar fi inutilă. Separarea datelor mai întâi se referă la crearea unei baze de date separate pentru datele pe care vrem să le folosim, acest tip de abordare aduce mai multe probleme la început întrucât ne face să ne gândim la probleme precum pierderea integrității datelor.

Atunci când facem modificări de tipul acesta trebuie să avem grijă cum orchestrăm schimbările. O modalitate ce provine de la un arbust tropical „Strangler Fig” se referă la crearea unui strat de interceptie ce verifică dacă funcționalitatea apelată a fost implementată într-un microserviciu, și dacă este atunci va ruta traficul către microserviciu. Acest tip de implementare permite să nu modificăm aplicația monolitică și să putem să ne întoarcem la ea în cazul în care avem probleme. De asemenea, dacă datele sunt comune putem folosi load balancing între cele două la nevoie. În rularea paralelă, microserviciul și aplicația monolitică funcționează în paralel și ne permite să le comparăm ca să ne asigurăm că avem funcționalitatea dorită. Metoda comutatorului ne permite să avem cele două

sisteme aplicația monolitică și microserviciul să funcționeze în același timp și să comutăm între cele două foarte ușor printr-un proxy.

În momentul în care începem separarea datelor, apar probleme ce nu pot fi prezise cauzate de lipsa unei baze de date monolitice. De exemplu, performanța unei baze de date este extrem de bună când este vorba de crearea de cereri ce conțin operații de alăturare de tabele, astfel noi trebuie să mutăm aceste operații extrem de optimizate în interiorul microserviciilor. O dată cu lipsa unei baze de date comună pentru toate tabelele apar problema lipsei de constrângeri, nu ne mai putem baza pe baza de date să forțeze aceste constrângeri și trebuie să verificăm și să tratăm erorile în cazul în care apar erori. Tranzacțiile nu mai sunt garantate. O dată cu separarea datelor se pierde ACID-itate, o altă problemă a microserviciilor atunci când avem tranzacții distribuite între mai multe microservicii. În funcție de necesități ar fi necesar să rulăm cereri direct în baza de date sau să oferim acces către persoane din exterior către aceasta, însă prin conceptul de ascundere de informație acest lucru nu ar fi posibil. Ca să rezolvăm această problemă putem crea o copie a bazei de date, read only, pe care microserviciul o va popula o dată ce populează baza de date inițială.

Astfel, migrarea unui sistem monolitic este extrem de dificilă și vor apărea probleme, din acest motiv este recomandat să folosim o abordare incrementală în care mutăm funcționalități una câte una, acoperindu-ne obiectivul pe care ni-l dorim prin introducerea arhitecturii formate din microservicii.

3.3 Tehnici de implementare

Trecerea de la o aplicație monolitică la una bazată pe microservicii uneori nu este ușoară. Unul din principalele motive este o posibilitate de a regândii modul cum accesăm anumite elemente. Atunci când totul se află în cadrul unui singur proces, accesarea datelor devine mai grea întrucât nu avem acces la ele și uneori nu este rentabil să căutăm după ele. Atunci când apelăm o funcție local, timpul necesar pentru executarea acesteia este neglijabil, atunci când această funcție trebuie să apeleze un alt microserviciu, lucrurile se complică și trebuie decis dacă chiar avem nevoie. Un apel făcut printr-o rețea mereu va avea o întârziere, de asemenea cresc locurile în care acesta poate eșua. În dezvoltarea aplicațiilor mobile întrucât ne așteptam ca aplicația să reacționeze la atingeri, nu putem să facem anumite operații pe thread-ul principal, acestea pot fi apelarea lucrurilor de pe Internet sau accesarea unei baze de date (chiar și dacă este de tipul SQLite3 fiind locală), aceasta pune în perspectivă impactul pe care acest tip de apeluri le pot avea.

Atunci când modificăm modul în care comunică un microserviciu trebuie să avem grijă la modul cum ar afecta alte microservicii, sau atunci când acest lucru este necesar trebuie să orchestrăm aplicarea acestor schimbări. Un alt concept ce într-o arhitectură monolitică se tratează ușor este tratarea erorilor, întrucât toate se întâmplă în același mediu de lucru

putem avea o viziune asupra ce poate să cadă și în ce mod. Într-un sistem distribuit, pot apărea mai multe tipuri de erori care se datorează rețelor, de exemplu crash-uri de microservicii ce necesită resetări, erori de omisie de mesaje, când încercăm să apelăm un microserviciu însă acesta nu trimite un răspund, în același timp acest răspund poate să fie mult prea rapid și să nu putem să îl recepționăm sau prea târziu și să considerăm că nu am primit, dar și erori arbitrare precum unele în care ne așteptăm la un răspund și nu primim ce avem nevoie.

Toate aceste aspecte trebuie tratate atunci când lucrăm cu microservicii, însă fiind vorba de o arhitectură paralelă ce oferă independență avem flexibilitatea de a alege ce folosim și ce nu. De aceea este important să alegem tehnologii și stiluri de comunicare care oferă performanțele dorite încercând să minizăm timpul de răspund și tipurile de avarii ce pot apărea.

3.3.1 Tipuri de comunicare

La fel ca în comunicarea între procese sau în cadrul unui proces, există comunicare sincronă ce introduce blocare la un anumit nivel și comunicare asincronă ce permite procesele să trateze cererile în momentul în care pot, aducând îmbunătățiri ale performanței dar crescând dificultatea înțelegerii și a modurilor în care aplicația poate cădea.

Cel mai simplu de înțeles mod de comunicare este comunicarea sincronă bazată pe cerere și răspund. În cadrul acestuia un microserviciu face o cerere (indiferent de formă) către un alt microserviciu și așteaptă un răspund. Acest scenariu este extrem de familiar, semănând cu cererile către o bază de date sau către un API extern, ceea ce le face ușor de implementat însă nu este o tactică eficientă în cadrul unui sistem distribuit întrucât aduc dezavantaje precum apariția acestui blocaj, creșterea latenței dar și apariția acestui cuplaj temporar întrucât se poate întâmpla ca unul dintre microservicii să nu fie același ca cel care a răspund întrucât a fost necesară o resetare. Acest tip de comunicare poate fi folosită pentru microservicii simple dar trebuie evitat crearea unui șir întreg de cereri întrucât crește timpul total de răspund dar și apariția unui eșec în cascadă.

Însă, o arhitectură paralelă are posibilitatea creării a unui tip de comunicare asincronă ce nu blochează funcționarea microserviciilor, adică ele nu așteaptă un răspund direct. Avantajul unui tip astfel tip de comunicare este posibilitatea tratării operațiunilor ce ar dura mai mult întrucât se elimină conceptul de cuplare temporară. Dezavantajul este creșterea nivelului de complexitate și creșterea dificultății de monitorizare și de a manevra erorile, însă ar trebui să fie utilizată atunci când avem procese de lungă durată sau pentru a comunica cu mai multe microservicii în același timp.

Cel mai simplu mod de a implementa comunicarea asincronă este crearea de resurse partajate. Un microserviciu adaugă date, fie într-o locație de stocare, adăugăm un fișier sau la o bază de date adăugăm alte date, iar celelalte microservicii ce vor să acționeze în

urma primirii acestei informații doar verifică apariția de date noi, iar atunci când apare o prelucrează. Dezavantaje includ necesitatea tratării cazurilor în care microserviciile ce scanează sunt căzute, iar în momentul în care vor reveni sunt nevoie să trateze date noi dar și faptul că se realizează o formă de cuplare asupra domeniului partajat. Această formă de comunicare poate fi folosită atunci când suntem restricționați de tehnologiile folosite, întrucât orice tip de sistem poate să citească date dintr-un fișier dar și atunci când vrem să trimitem un volum de date ridicat.

Un alt tip de comunicare este cel sub formă cerere-răspuns, acesta poate fi sincron sau asincron. În mod blocant, microserviciul ce face cererea va aștepta un răspuns iar în funcție de acesta își va continua activitatea. În mod asincron, se va proceda în mod asemănător însă de data aceasta nu se va mai aștepta un răspuns, sau cel puțin nu unul cu informația cerută. Acest lucru este mai dificil, întrucât pentru a răspunde cererii, microserviciul trebuie să trimită unui microserviciu specific răspunsul creat și există posibilitatea ca microserviciul ce inițiază cererea să nu mai existe astfel am putea încerca ca orice microserviciu să poată să trateze orice cerere. Putem folosi acest mod de implementare atunci când avem nevoie să avem o confirmare la cererile făcute și de asemenea este destul de ușor de implementat, cel puțin varianta blocantă.

Una dintre comunicările reprezentative pentru microservicii este cea făcută prin evenimente, este asincronă în natură și se bazează că atunci când în cadrul unui microserviciu se întâmplă un lucru pe care consideră că ar trebui să îl comunice în exterior, acesta emite un eveniment cu diferite informații astfel celelalte microservicii pot asculta iar în momentul primirii acestea vor reacționa independent. Este un opus al modelului anterior întrucât în acest tip de comunicare, în general, nu avem cunoștințe dacă evenimentul a ajuns către celelalte părți sau dacă cineva ascultă pentru acest tip de evenimente ci eventual dacă a fost procesat de către Event Bus. Comunicarea prin evenimente oglindește modul de organizare al unor echipe autonome din cadrul companiilor, acestea anunță atunci când produc un rezultat nou și nu primesc feedback direct dacă s-a acționat după mesajul lor în mod implicit, întrucât comunicarea poate fi un mail general la o listă de abonați, aceștia ar primi doar confirmarea că serverul ce procesează mail-urile a validat cererea. Implementarea constă în crearea unei metode de a trimite evenimente și de a recepționa. Un prim mod este folosirea de programe specializate precum Apache Kafka sau RabbitMQ ce livrează platforme dar și implementări pentru ambele probleme ce vin cu avantajul că toate problemele care le-am avea precum garantarea livrării și permit scalare însă sunt un sistem în plus de administrat. Un alt mod de implementare include folosirea specificațiilor declarate în Atom pentru a crea un flux web de resurse ce poate fi folosit pentru a trimite și consuma evenimente însă nu oferă latență mică. În general, ar fi bine să nu implementăm comunicarea prin evenimente de la zero întrucât este dificil de administrat și sunt multe probleme greu de tratat, de aceea putem folosi produse existente. Avantajul unui astfel de tip de comunicare este posibilitatea ca un eveniment să

fie tratat de mai multe microservicii în același timp, astfel evenimentul trebuie să conțină informații suplimentare pentru ambele microservicii, deci apare riscul de a distribui mai multe informații ceea ce induce diferite probleme de securitate, astfel putem folosi acest tip de arhitectură acolo unde putem profita de acest avantaj.

Deși comunicarea asincronă aduce îmbunătățiri în decuplarea sistemelor și în timpul de răspuns în cazul în care vrem să comunicăm cu mai multe evenimente acesta aduce complexități ce devin greu de depanat, astfel avem nevoie de monitorizare și moduri de identificare a modului evoluției cererii implementate corespunzător.

3.3.2 Tehnologii de comunicare

Am menționat anterior că microserviciile ne cumpără flexibilitate, însă din acest motiv avem și mai multe probleme întrucât mereu când alegem un tip de tehnologie trebuie să limităm dezavantajele și în același timp să maximizăm avantajele pe care le poate aduce. Astfel putem să alegem tehnologii mai vechi care au fost testate dar limitate sau tehnologii noi care promit că aduc îmbunătățiri dar e posibil să nu fie lipsite de erori ce s-ar rezolva în timp.

Când alegem o tehnologie, trebuie să avem grijă că aceasta trebuie să fie compatibilită cu serviciile pe care le avem, de asemenea trebuie să nu ne limiteze în alegerea tehnologiilor viitoare. În acest scop modul de comunicare al microserviciilor ar trebui să fie clar definit întrucât nu putem fi siguri ce elemente sunt folosite și care nu. De asemenea tehnologia aleasă nu ar trebui să inhibe comunicarea cu alte microservicii.

Remote procedural call (RPC) (Apel procedural la distanță) este un mod de apelare de funcții local însă care se execută într-un serviciu la distanță. Pentru a face asta posibil sunt necesare definirea unor scheme explicite precum SOAP sau gRPC definite ca limbaje de definit interfețe (interface definition language). Definirea unor scheme explicite face ca folosirea tehnologiilor diferite să fie mai ușoară însă dacă nu vrem să folosim astfel de scheme putem folosi Java RMI care necesită folosirea aceleiași tehnologii însă fără schemă definită. Folosind o astfel de tehnologie ne permite generare de cod pe care clientul le poate folosi pentru a comunica cu serverul făcând implementarea pe partea clientului mai ușoară întrucât nu trebuie să facă ceva în mod special. Limitările acestei tehnologii reprezintă restricționarea tehnologiilor folosite (mai ales în cazul Java RMI) însă celelalte menționate funcționează cu diferite tehnologii. Ascunderea faptului că prin aceste metode se face un apel peste Internet poate fi înșelător și ascunde complexitatea. De asemenea atunci când producem modificări asupra codului serverului trebuie să generăm codul pentru client, însă dacă adăugăm sau eliminăm câmpuri din interiorul funcțiilor, clientul trebuie să facă mai multe modificări deși el poate să nu folosească acele părți din cod. Acest tip de abordare este potrivit pentru comunicarea blocantă între două sisteme asupra cărora știm când se modifică lucrurile și în ce fel, astfel cunoaștem cât de multe modificări vor trebui

să fie făcute.

Representational State Transfer (REST) prin HTTP, în acest tip de arhitectură avem definite niște resurse ce pot fi accesate prin link-uri sub diferite forme ce permite identificarea și modificarea acestora în funcție de modul de accesare. Acest tip de abordare funcționează foarte bine cu verbele HTTP predefinite (GET, POST, UPDATE) întrucât știm la ce comportament ne putem aștepta indiferent de resursa pe care o accesăm. Faptul că folosim HTTP ne permite să folosim toate elementele dezvoltate pentru a îmbunătății sistemul precum caching, echilibrări de trafic, unelte de monitorizare existente dar și elemente de securitate precum autentificare și certificate. O altă trăsătură adusă în REST este Hypermedia as the engine of application state (HATEOS) ce permite să nu cunoaștem întreg URI-ul către resursă ci doar cel de început. O dată intrat pe prima intrare, aceasta va afișa alte intrări pe care putem intra și ce am putea accesa dacă intrăm pe ele. Acest lucru este util pentru că nu trebuie să definim niște căi explicite ci acestea pot fi căutate însă pentru a face asta cresc numărul de cereri pe care trebuie să le faci doar pentru a executa o singură comandă. Atunci când folosim această tehnologie nu avem o modalitate simplă de a crea librării pe care clienții serverului le pot folosi, din acest motiv s-a creat specificația OpenAPI din Swagger ce permite ca din anumite informații să generezi automat o documentație dar și cod pentru o varietate de limbaje însă nu a fost adoptat la fel de mult întrucât ar fi dificil de implementat pentru proiectele existente. Prin acest format putem trimite o varietate de formate de date însă din cauza cerințelor adăugate de HTTP atunci când creăm o cerere ar putea să împiedice performanțe ridicate atunci când nu avem date multe de trimis. Putem folosi acest mod de comunicare cam în toate situațiile în care avem nevoie de comunicare unu la unu între servicii, fiind un stil comun de comunicare este suportat de majoritatea tehnologiilor deci ne garantează ca am avea compatibilitate.

Ambele tipuri de apeluri au aceeași problemă, în unele cazuri ar trebui să facem mai multe apeluri pentru aceeași resursă întrucât nu avem suficiente date, sau în cazul în care un câmp dintr-o cerere se schimbă acesta ne-ar forța să modificăm în mai multe locuri. Astfel de probleme sunt rezolvate de GraphQL ce permite să specificăm exact ce câmpuri avem nevoie fără a fi nevoie să facem mai multe cereri. Deși ar ușura cererile la nivelul clientului, dezavantajele apar la server, în primul rând acesta este un Query Language, deci dacă cererea este complicată atunci server-ul va trebui să o proceseze, dacă se adună multe cereri de tipul acesta ar scădea performanța. De asemenea, întrucât de fiecare dată se face un query, caching-ul este mai dificil de realizat și nu poate fi folosit la fel de eficient pentru a scri informații.

În arhitectura bazată pe evenimente putem folosi agenți de mesaje (Message Brokers), aceștia funcționează ca un intermediar pentru comunicațiile între microservicii, astfel în loc să comunice direct acestea comunică prin intermediar trimițând orice tip de mesaje. Acestea oferă diferite facilități în modul în care trimit mesaje, printr-o coadă în care se

pun mesajele iar consumatorii doar preiau direct din coada aceasta sau prin topică ce permite consumatorilor să se aboneze la un anumit tip de mesaj iar atunci când un mesaj vine ce îndeplinește criteriul, acesta va fi trimis la toți abonații. Principalul motiv pentru care am folosi o astfel de tehnologie este pentru că oferă diferite caracteristici care ar fi greu de implementat precum garanția livrării, astfel dacă nu există nici-un microserviciu care să accepte un mesaj, atunci el va fi livrat la primul care va fi disponibil să prelucreze acel mesaj. Alte caracteristici includ garantarea ordinei în care mesajele sunt primite sau să îți permită să trimiți același mesaj către mai multe topic.

Asupra acestor diferite moduri de comunicare apar diferite moduri de a trimite un mesaj, referindu-ne la compoziția acestora. Astfel, acestea pot fi sub formă de text sau sub formă binară. În general, putem trimite sub formă text adică JSON sau XML însă dacă o să ne intereseze să eficientizăm viteza atunci va fi necesar să encodăm mesajele trimise și să le trimitem sub formă binară însă acesta nu este un proces ce ar aduce suficient de multe îmbunătățiri.

Atunci când am vorbit de comunicarea între microservicii am menționat importanța creării unei interfețe sau a unei scheme pentru a afișa modul cum comunică microsericiul cu exteriorul. Acest lucru permite consumatorilor să știe la ce să se aștepte și să implementeze mult mai ușor crearea legăturii iar pentru cei ce administrează microserviciul ar cunoaște ce câmpuri sunt expuse și ce trebuie să facă. Deși unele tehnologii necesită prezența unei scheme sub o formă sau alta, schemele ne ajută și în crearea unei documentații consistente ce ajută la implementare. Cu ajutorul unei scheme putem găsi probleme mult mai ușor, asemănător cu modul în care funcționează limbajele de programare ce necesită specificarea unui tip atunci când declarăm variabile și limbajele interpretate ce pot avea orice valoare, avem un nivel în plus de siguranță, cel la nivelul la compilare, astfel într-un limbaj în care acest tip de verificare ar lipsii, unele probleme apar doar în producție. Atunci când trebuie să modificăm o schemă aceasta se poate face în doua feluri, structural când câmpurile se schimbă, astfel clienții trebuie sunt nevoiți să administreze noul câmp, sau semantice atunci când câmpul nu își schimbă numele însă modul de calculare este diferit. Acest ultim tip este destul de periculos întrucât este ușor de ignorat dacă nu apar probleme evidente. În general este recomandat să creăm scheme de tipul acesta, mai ales dacă microserviciile sunt deținute de echipe diferite, în cazul acesta nivelul de comunicare dintre cele două este îmbunătățit iar problemele sunt rezolvate mult mai ușor.

3.3.3 Schimbări în funcționalitate

Atunci când este vorba de microservicii, este inevitabil ca funcționalitatea acestora să se schimbe la un moment dat, iar în cazul cel mai fericit în care cuplarea și coeziunea sunt favorabile, atunci cel mai probabil nu trebuie să modificăm alte microservicii, însă de multe ori acesta nu este cazul.

O primă parte sunt metodele de evitare ale acestor probleme, cel mai simplu prin asigurarea că funcționalitatea microserviciului nu este schimbată ci doar se adaugă funcționalitate. De asemenea, putem implementa în consumatori ca aceștia să fie mai toleranți și să încerce să își găsească datele de care au nevoie chiar dacă se schimbă schema, acest lucru se poate face și prin prelucrarea strictă a câmpurilor de care avem nevoie și nu de întreg fișierul. Putem folosi tehnologii care să permită adaptare ușoară, de exemplu atunci când serializăm un obiect în Java acesta are o anumită versiune, însă dacă modificăm obiectul adăugând un câmp această versiune s-ar schimba deși câmpurile de care am avea nevoie ar rămâne la fel, acest lucru creând o incompatibilitate, însă dacă am folosi ceva asemănător cu HATEOS ar fi mult mai ușor de adaptat fără nicio modificare. Este important ca orice modificare adusă datorită unei schimbări în funcționalitatea unui microserviciu să fie prinsă cât mai rapid, de asta sunt importante testele iar cu ajutorul schemelor definite explicit putem folosi diferite unelte ce pot valida diferențe între modificări la nivelul protoalelor, schemelor sau a specificațiilor OpenAPI.

Însă uneori este imposibil să mitigăm toate problemele și este nevoie să adăugăm schimbări pe care și alte microservicii sunt obligate să le facă, astfel apar dificultăți în orchestrarea ordinii acestor schimbări. Una dintre soluții este anunțarea schimbărilor pe care vrem să le facem și să lăsăm toți consumatorii să lanseze versiuni noi ce tolerează schimbările și după aceea să lansăm varianta schimbată. Acest lucru nu este în general suficient întrucât datorită modului de planificare aceste apariții acestor schimbări în consumatori poate să fie destul de lentă. O altă metodă de administrare a schimbărilor este de a lansa două servicii în paralel, una ce conține modificările și una care nu le conține pentru microserviciile care încă nu s-au adaptat. Acest lucru este de evitat din același motiv ca precedentul dar și faptul că crează multe dificultăți să ai mai multe servere de același timp lansate. De asemenea se poate păstra versiunea veche și doar crearea unei noi versiuni a endpoint-ului. Versiunile mai vechi pot fi schimbate să fie compatibile cu versiunea nouă și eventual retrase când suntem siguri că toți clienții au primit modificările necesare. Acest lucru se poate face prin monitorizarea traficului pe interfață. Acesta este tipul preferat de adus schimbări întrucât este cel mai convenabil pentru consumatori dar și pentru server.

3.3.4 Partajare între microservicii

Unul dintre cele mai simple concepte de programare ce aduce îmbunătățiri în scrierea codului este DRY, „Să nu te repeți” (Don't Repeat Yourself) acesta la prima vedere are ca sens să nu duplici codul întrucât dacă ar trebui să schimbi o funcționalitate e posibil să fie nevoie să modifice în mai multe locuri din cadrul sistemului ceea ce face ca modificările să fie mai greu de făcut.

În universul microserviciilor, acestea sunt separate din natură, astfel repetarea codului

în cadrul mai multor microservicii este inevitabilă, de asta am fi tentați să creăm librării comune (de exemplu pachete de NPM sau dependențe în Maven) însă chiar dacă inițial ar fi mai ușor de scris acest lucru generează o cuplare ridicată între aceste sisteme deoarece atunci când o librărie comună este actualizată atunci sistemele trebuie să implementeze versiunea nouă, însă nu de multe ori este posibil acest lucru întrucât nu toate microserviciile ar putea fi actualizate în același timp, în acest scop trebuie să înțelegem că librăria creată de noi va fi consumată sub forma unor versiuni multiple de clienți diferiți.

Un mod avansat de a folosi diferite microservicii este crearea unei librării pentru acel microserviciu pe care mai târziu o oferim altor consumatori. Când este vorba de echipe mici, este dificil de implementat întrucât logica de server ar trebui să fie diferită de logica ce se găsește în librăria de client, când librăria ar trebui să funcționeze ca o abstractizare al API-ului pe care îl oferă microserviciul, astfel e indicat ca cele două să fie separate. La fel ca librăriile trecute trebuie să ne obișnuim cu ideea că librăria va fi consumată în mai multe versiuni.

Când lansăm microservicii trebuie să ne asigurăm că acestea pot să comunice între ele. Acest lucru, în general, este dificil de implementat întrucât microserviciile pot rula pe servere care dispar și apar din nou constant, mai ales când e vorba de infrastructură oferită ca serviciu. Sisteme de orchestrare precum Kubernetes ne oferă o metodă simplă de configurare a microserviciilor ce se află într-un singur cluster însă atunci când introducem mai multe, devine mai complicat. Însă chiar și așa nu ne putem baza pe serviciul de orchestrare să ne facă legătura întrucât dacă avem un număr mic de microservicii nu ar fi suficient de rentabil să rulăm o astfel de tehnologie.

O metodă simplă folosită și pentru Internet este Domain Name System (Sistem de nume de domeniu) prescurtat DNS. Acesta ne permite să asociem unui IP un nume, astfel nu ar trebui să cunoaștem exact IP-ul sistemului cu care vrem să comunicăm ci doar numele acestuia. Avantajele acestei metode este familiaritatea tuturor persoanelor cu acest lucru, din acest motiv ansamblarea unui sistem nu ar trebui să fie grea, însă în cadrul microserviciilor apar mai multe probleme datorită faptului că acestea pot să fie înlocuite mereu de instanțe noi. Fiecare intrare de DNS are o perioadă în care nu este actualizată (Time to live - TTL), și poate fi cached sub diferite forme. Din acest motiv putem opta pentru varianta în care păstrăm în DNS doar calea către un load-balancer sub care se află microserviciile de care avem nevoie, însă această soluție aduce problema în care nu știm exact cu cine comunicăm iar dacă unul din ele are probleme acesta nu va fi evitat.

Limitările DNS-ului au adus la apariția tehnologiilor sub nume de Registrii dinamice de servicii Dynamic Service Registries ce presupune ca microserviciul să fie cel care se înregistrează la un serviciu central iar pentru ceilalți ce vor să îl acceseze vor lua intrarea din cadrul acestui centru. Astfel de sisteme includ ZooKeeper ce necesită rularea mai multor noduri într-un cluster ce poate funcționa ca un sistem de configurare și permite

crearea unei ierarhii în care se pot insera, modifica și interogă noduri sub diferite forme. Consul este asemănător cu tehnologia precedentă însă are mai multe facilități printre care include o interfață RESTful prin HTTP ce face integrarea cu diferite tipuri de tehnologii mult mai ușoară sau prin faptul că se pot genera fișiere de configurare ce vor fi actualizate în timp real, ceea ce permite ca programul să nu interacționeze cu Consul. O altă metodă am menționat-o anterior, adică prin Kubernetes, modul în care lucrează acesta este că prin faptul că microserviciile sunt lansate prin aceasta platformă, aceasta poate să înregistreze modul în care a evoluat și să își păstreze un serviciu ce poate identifica locația pod-urilor lansate prin el. Kubernetes oferă multe servicii fără nicio configurare ce permite să evităm platforme dedicate ca cele două menționate anterior.

Serviciile Cloud în general se împart în regiuni ce oferă servicii. În cadrul acelor regiuni putem avea mai multe resurse ce ar fi grupate sub același centru. Putem defini ca trafic Nord-Sud, traficul ce se află în afara centrului de date (de exemplu clienți) și Vest-Est traficul ce se produce în interior.

Pentru a facilita comunicarea Nord-Sud există conceptul de API Gateway (Pasarelă API) ce funcționează ca un intermediar între persoanele din exterior ce vor să acceseze microserviciile din interior. Un segment de piață care s-a dezvoltat recent este acela de a monetiza API-uri create prin diverse metode precum abonamente cu un număr limitat de cereri. Când vrem să oferim un astfel de serviciu apariția unui intermediar ce poate să îndeplinească diferite facilități precum monitorizare, limitare de trafic sau funcționarea ca un reverse proxy și poate să ofere un fel de panou de control, însă în general este folosit de către companii pentru a-și accesa microserviciile din aplicațiile dezvoltate, acest lucru se întâmplă pentru că în general un API Gateway este destul de complex și uneori nu avem nevoie de toată funcționalitatea pe care ne-o oferă, de asemenea pot să fi și alte soluții care să fie mai eficiente. De exemplu, dacă vrem să externalizăm microservicii ce rulează în Kubernetes putem folosi soluții precum Ambassador. Dacă vrem un API Gateway întrucât ne așteptăm la mult trafic, atunci probabil nu avem nevoie de ceva central ci de mai multe. Un API Gateway poate funcționa însă trebuie să știm dacă se pliază cu funcționalitatea de care avem nevoie și să folosim alte tehnologii mai specializate dacă nu găsim.

Comunicarea Vest-Est se poate îmbunătăți folosind Service Meshes, aceasta îmbunătățește comunicarea între microservicii prin adăugarea unui proxy, astfel microserviciile nu ar comunica direct ci fiecare ar avea propriul proxy cu care comunică iar acestea vor comunica între ele. Ele sunt administrare de servicii de control și îmbunătățește traficul întrucât proxy-urile pot fi în aceeași rețea locală. Acest lucru simplifică dezvoltarea întrucât microserviciile doar fac cereri normale însă de fapt sunt rutate de către proxy. Deși utile, acestea induc multă complexitate în infrastructură ceea ce e bine să fie evitată.

Îmbunătățirea metodelor de comunicare între microservicii nu este suficientă pentru a crea un sistem complet pentru care este ușor să dezvoltăm. Am menționat importanța

schemelor explicite, acestea au ca rol îmbunătățirea comunicării capabilităților unui unct de acces prin prezentarea schemei însă nu este suficient, de asta e necesară documentație explicită. Standardul OpenAPI este eficient în crearea eficientă a documentației combinând cu Ambassador pentru Kubernetes poate să descopere automat puncte de acces OpenAPI și să afișeze documentația automat. Documentația pentru comunicarea bazată pe evenimente nu este la fel de dezvoltată însă avem opțiuni precum AsyncAPI sau CloudEvents.

Atunci când avem multe sisteme devine din ce în ce mai greu să cunoaștem exact ce face fiecare sistem, de asta e important să știm starea fiecăruia. Folosind ID-uri de corelare pentru tranzacții putem să vedem modul cum interacționează. Folosind servicii de descoperire precum Consul putem să vedem unde rulează microserviciile, OpenAPI ne arată ce capabilități are fiecare sistem și sistemele de monitorizare ne arată starea sistemelor. Toate informațiile acestea sunt disponibile prin intermediul diferitelor API-urilor de care dispunem atunci când folosim aceste tehnologii. De asta este destul de important să avem pagini dedicate care se actualizează în timp real ce arată nu doar documentația microserviciilor ci și starea acestora. O astfel de unealtă a fost construită la Financial Times, Biz Ops, ce oferă starea întregului sistem și calculează pentru fiecare o metrică ce ar indica dacă acestea au nevoie de schimbări. Astfel de unelte există disponibile și liber, precum Backstage de la Spotify.

O arhitectură distribuită ne permite creșterea eficienței prin distribuirea sarcinilor către mai multe sisteme de calcul. Într-o aplicație monolitică putem să obținem performanțe mai bune prin eficientizarea algoritmilor, prin folosirea thread-urilor în interiorului sau prin crearea unor procese copil, însă toate acestea rulează sub un singur procesor, ceea ce înseamnă că indiferent că se întâmplă aproximativ concurrent, nu este chiar în totalitate. Însă cu ajutorul microserviciilor, acest lucru este posibil.

Însă atunci când apare un lanț lung de instrucțiuni, înseamnă că sunt multe puncte în care acesta se poate întrerupe și să aibă nevoie să se facă o analiză a lucrurilor care s-au întâmplat și să se întoarcă la o versiune sigură. Acest concept apare în contextul prezenței unei baze de date, în general relaționale însă există și versiuni care nu sunt relaționale care promit același lucru. O tranzacție reprezintă un grup de instrucțiuni ce ar trebui să se execute împreună, dacă una din ele eșuează din diverse motive, atunci toate acestea sunt anulate și ne întoarcem la un punct de recuperare de la începutul tranzacției. Acest lucru se referă la „aciditatea” bazei de date ce fac parte din acronimul „ACID” ce reprezintă un set de proprietăți ce se referă la consistența datelor dintr-o bază de date. Consistența se referă că în cadrul unei operațiuni cu baza de date, la finalul acesteia este lăsată într-o stare validă pe care se poate continua să se execute operațiuni. Izolarea se referă ca multiple tranzacții se pot întâmpla în același timp fără ca cealaltă să fie afectată, acest lucru este asigurat de separare stărilor intermediare către o altă tranzacție. Durabilitatea se referă că la finalul unei tranzacții, datele nu sunt pierdute dacă sistemul cade.

În cadrul unui sistem format din microservicii, putem avea tranazații de tipul acesta, în general atunci când microserviciul interacționează cu datele din baza de date proprie, iar microserviciul are control complet asupra acestora. Problemele apar atunci când modificările trebuie să fie făcute în cadrul mai multor microservicii în același timp iar asupra acestora să rămână proprietățile ce oferă consistența datelor. Acesta este un lucru mai dificil de implementat și e nevoie de mai mulți pași pentru a avea un comportament asemănător ca în cadrul unui sistem monolitic.

3.4 Build

3.5 Testare

3.6 Deployment

3.7 Monitorizare

3.8 Securizare

3.9 Evolutie

Capitolul 4

Orchestrare

4.1 Istoric deployment

4.2 Tipuri de deployment

4.2.1 Baremetal

4.2.2 Mașină virtuală

4.2.3 Container

4.2.4 Function as a Service

4.2.5 Platform as a Service

4.2.6 Container as a Service

4.3 Docker

4.4 Docker Swarm

4.5 Kubernetes

4.6 Red Hat OpenShift

4.7 HashiCorp Nomad

4.8 Infrastructure as a Service

Capitolul 5

Îmbunătățirea metodelor de dezvoltare software

5.1 Schimbari in livrarea produselor

5.2 DevOps și aspecte ale comunitații

5.3 Ce înseamnă să aplici DevOps

5.4 Unelte

Capitolul 6

Aplicație

Capitolul 7

Concluzii

Bibliografie

- [1] Bahrim Dragoș, „Temă „Principii SOLID și CI/CD în Microservicii” pentru cursul Metode de dezvoltare software”, în (Iunie, 2022).
- [2] Albert Endres și H. Dieter Rombach, *A Handbook of Software and Systems Engineering: Empirical observations, laws, and theories*, Pearson/Addison Wesley, 2003.
- [3] Eric Evans, *Domain-Driven Design: Tackling Complexity in the Heart of Software*, Addison-Wesley Professional, August 20, 2003.
- [4] David Lorge Parnas, „Information Distribution Aspects of Design Methodology”, în *Information Processing, Proceedings of IFIP Congress 1971, Volume 1 - Foundations and Systems, Ljubljana, Yugoslavia, August 23-28, 1971*, ed. de Charles V. Freiman, John E. Griffith și Jack L. Rosenfeld, North-Holland, 1971.