

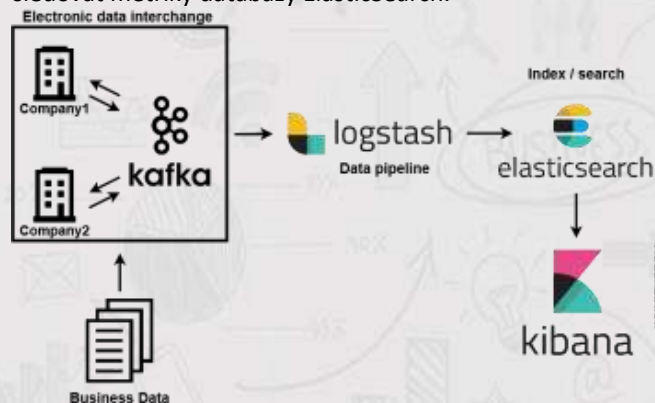
Motivácia

Ľudia vo všeobecnosti obľubujú zjednodušovanie a zefektívňovanie práce, a preto sa kladie veľký dôraz na vývoj technológií, ktoré tieto faktory zabezpečujú. Obchodný sektor je jedným z týchto faktorov. Elektronická výmena obchodných dokumentov je ekonomickejšia aj ekologickejšia oproti papierovej. V dnešnej dobe môžu tento proces uľahčiť aj technológie Veľkých dát. Tieto technológie zabezpečujú to, že je možné pracovať s obrovským množstvom dát rôzneho druhu s veľkou rýchlosťou. Cieľom tejto záverečnej práce je navrhnúť architektúru na spracovanie a analýzu transkačných obchodných dát s využitím technológií Veľkých dát. Cieľom tejto práce bolo aj naimplementovať program fungujúci na princípe výmeny dokumentov medzi firmami v systéme. Dáta, ktoré si medzi sebou firmy posielali boli faktúry v štandarde Universal Business Language.

Riešenie

Prvým krokom je načítanie obchodných dokumentov, faktúr, zo zdrojovej databázy súborového systému. Po skonvertovaní faktúr na JSON formát môžeme začať s implementáciou Kafka klienta. Implementácia klienta je navrhnutá na princípe firiem, ktorej súčasťou sú produceri a consumeri. Tí nám umožňujú výmenu faktúr prostredníctvom Kafka. V tomto kroku pracujeme s faktúrami vo formáte JSON a cez Kafku tieto dokumenty odosielame a prijímame. Princípom tejto architektúry je správne nastavenie klienta Apache Kafka, ktorý bude zabezpečovať prenos obchodných dokumentov medzi producermi a consumermi v určitých časových intervaloch. Dôležité je vytvoriť si topic, prostredníctvom ktorého bude Kafka vedieť, ktorému consumerovi má obchodné dokumenty poslať. Následne je nevyhnutné zabezpečiť to, aby sa tieto dokumenty aj ukladali do databázy. Databázu, ktorú sme v tomto prípade použili sa volá ElasticSearch. ElasticSearch sme prepojili s Kafkou cez dátový kanál Logstash prostredníctvom konfiguračného súboru dátového kanálu. V tomto kroku tento dátový kanál má aj zaindexovať obchodné dokumentov, aby sme následne vedeli pracovať ďalej s týmito dátami.

Konfiguračný súbor je upravený tak, že na vstupe bude mať správy z Kafky, ktoré sa budú ukladať do ElasticSearchu. Je nevyhnutné aby sme si vedeli dáta, ktoré sú uložené v Elasticu zobraziť v určitom rozhraní. Na tento konkrétny prípad sme použili webové rozhranie Kibana, kde sme sledovali jednotlivé operácie vykonané nad dátami. V tomto rozhraní vieme tiež sledovať metriky databázy ElasticSearch.



Testovanie a výsledky

Na testovanie sme naimplementovali program na simulovanie komunikácie medzi firmami. Tento program má na vstupe 3 parametre – počet firiem, počet obchodných partnerov a počet dokumentov, ktoré jedna firma posielala druhej. Tieto parametre určujú to ako veľmi sa v systéme komunikuje. Následne sme vytvorili tri testovacie scenáre, kde sme škálovali vstupné parametre. Program meral dĺžku celkového trvania komunikácie, ktoré pozostáva z dĺžky posielania, dĺžky prijímania a dĺžky odpovedania. Testovanie bolo vykonávané na školských serveroch, kde na jednom serveri bola spustená navrhnutá architektúra a na 3 serveroch boli spustený naimplementovaní klienti pre jednotlivé testovacie scenáre.

V prvom testovacom scenári bolo v systéme vytvorených 50 firiem, v druhom 75 firiem a v treťom testovacom scenári 100 firiem. Parameter počtu obchodných partnerov menil svoje hodnoty s následnou postupnosťou – 15, 25, 25, 35, 35. Parameter počtu odoslaných dokumentov menil svoje hodnoty s touto postupnosťou – 25, 25, 35, 25, 35.

Dĺžka komunikácie

	1.testovací scenár	2.testovací scenár	3.testovací scenár
1.prípád	489,12	675,87	983,96
2.prípád	588,51	1040,74	1694,98
3.prípád	807,84	1876,89	3331,21
4.prípád	954,79	2071,91	3558,20
5.prípád	1474,45	2690,25	5668,20

Po vykonaní všetkých troch testovacích scenárov môžeme konštatovať, že ak zvyšujeme hodnoty jednotlivých parametrov, komunikácia trvá dlhšie. Na základe výsledkov môžeme takisto skonštatovať, že odosielanie obchodných dokumentov trvá dlhšie ako prijímanie týchto dokumentov. Pri výpočtovej sile, ktorú sme mali k dispozícii, má najväčší vplyv na výkon systému tretí parameter, ktorý určuje koľko správ má odoslať jedna firma každej firme zo svojho zoznamu obchodných partnerov. Tento parameter výrazne ovplyvňuje odpoveď na prijaté obchodné dokumenty. Výsledky však môžu byť iné ak by sme jednotlivých klientov spúšťali na zariadeniach, ktoré by mali výpočtovú silu väčšiu alebo na druhej strane, menšiu ako sme mali k dispozícii.

Zhrnutie

Odporúčaním do budúcnosti v oblasti tejto práce je návrh zmeny dátového kanála, keďže dátový kanál Logstash spôsobuje veľkú záťaž na pamäť RAM. V tomto prípade by mohol byť využitý konektor Kafka Connect, ktorý nie len prepája rôzne služby, ale môže v sebe obsahovať aj serializér a deserializér na využívané formáty. Navrhnutá architektúra by mohla byť taktiež implementovaná v systéme komunikácie dodávateľov a odberateľov, ktorý od nich dané produkty odoberajú, na to aby sa zefektívnila ich komunikácia a aby dodávateľia mali k dispozícii jednoduchší proces analyzovania nákupov svojich zákazníkov.