

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Architektura systému pro analýzu online trhu
Jméno autora:	Otto Vodvářka
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Oponent práce:	Ing. Karel Frajták, PhD.
Pracoviště oponenta práce:	Katedra počítačů

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	mimořádně náročné
Zadání je spíše náročnější. Autor musel napřed provést analýzu problémů stávajícího řešení a navrhnout novou lepší a výkonnější verzi.	

Splnění zadání	splněno s menšími výhradami
Zadání bylo splněno. Výhrady jsou uvedeny níže.	

Zvolený postup řešení	správný
Zvolená metoda řešení je správná. Nejdříve byly identifikovány problémy stávajícího systému, následně stanoveny požadavky na systém nový a ten následně implementován. Oceňuji definici detailních testovacích scénářů. Jak zmiňuji níže, související kapitoly jsou od sebe poměrně daleko, zde by reorganizace textu pomohla – definoval by se problém, přidal popis technologií a řešení, stanovil testovací scénář a jeho výsledek. V rešerši není zmíněna actor-based architektura.	

Odborná úroveň	B - velmi dobře
Student využil znalosti získané studiem.	

Formální a jazyková úroveň, rozsah práce	C - dobře
Po formální stránce nejde práci nic vytknout. Práce je ale psána poměrně složitým jazykem s častými překlepy, které snižují její váhu. Typografie práce je v pořádku. Ač někdy mám výtky k rozsahu práce, tato práce je obsáhlejší, ale některé pasáže jsou delší než by měly a jiné zase dost krátké. Práce by si zasloužila zredigovat. Některé související kapitoly jsou od sebe až moc vzdálené – mikroslužby, database per service, 12 faktorové aplikace, observability, sidecar, ...	

Výběr zdrojů, korektnost citací	B - velmi dobře
Zdroje odkazují na stránky na internetu týkající se dané technologické problematiky. Citace jsou úplně a v souladu s normami. Do rešerše by ale bylo vhodné zahrnout i nějaké vědecké publikace.	

Další komentáře a hodnocení
Autor navrhuje řešení pro těžení dat z internetu z různých zdrojů pro jejich další analýzu. Bohužel toto není dále rozvedeno z důvodů NDA se zadavatelem a je ponecháno čtenáři bez dalšího rozvedení. Hned na začátku je u jedné z potřeb uživatele uveden příklad „Co dělají úspěšné weby?“, pokud je to ve stejné oblasti, pak to dává smysl, jinak samozřejmě ne. Autor uvádí analýzu založenou na klíčových slovech, návštěvnosti a proklících z reklam, ale neuvádí, jak se k takovým datům dostat bez přístupu k nástroji Google Analytics apod., což není zmíněno. Často je v textu zmíněn nějaký termín, který je vysvětlen až později nebo vůbec – pod, K8s operátor V textu je řada tvrzení, která jsou sporná nebo nepodložená

- monolitický systém je obtížný a drahý na vývoj (neplatí vždy), jeho vrstvy lze podle autora nasazovat separátně (jak?).
- SOA není obtížná na vývoj.
- Event-driven architektura se soustředí na vysokou škálovatelnost (to není priorita). "
- Problém sdílené databáze není jen single-point-of-failure. "
- V event-sourcingu je možné odebírat události.
- Mezi 3 velmi populární styly komunikace je zařazen mechanismus service discovery.
- Serverless architektura neznamena, že celý systém je nasazený v cloudu, ale v definici systému pomocí funkcí, které už dnes mohou být stavové.
- Quarkus určitě zdokumentován bude.
- Co jsou to obecně prováděcí grafy?
- Logování je implementací vzoru.
- LogStash filtruje a transformuje data, která jsou nakonec nestrukturovaná (viz screenshoty z Kibany).
- Bod XI pro 12F není splněn právě použitím ELK stacku.

Nebo tvrzení bez o problémech bez možných návrhů řešení

- event-sourcing a transakční odeslání zprávy (transactional oubox);
- běžné problémy Springu;
- Vert.X má podobně jako Spring cloud řadu funkcionalit (jakých?);
- metadata u MicroNautu (jaká?)

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Autor vystavěl systém pro těžení a zpracování velkého množství dat pomocí poměrně komplikovaných technologií. Svoji práci mohl prezentovat lépe. Další komentáře kromě výše zmíněných nemám.

1. Systémem protéká velké množství dat. Přesto jste zvolil dávkové zpracování. Proč nebylo zvoleno proudové (streamy) zpracování?
2. Proč je na každém podu přítomen Filebeat pro sběr logů, když jinde v textu píšete, že aplikace nemá nijak řešit logování a logy se zapisují na standardní výstup a ne do souboru?
3. Proč je v systému druhý systém pro vizualizaci dat (Grafana) a alternativní nástroj pro sběr metric (Prometheus), když šlo použít Metricbeat?
4. Byly testy nějakým způsobem automatizovány? Jak jste ověřoval jejich výsledky?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **C - dobře**.

Datum: 10.6.2022

Podpis: