MKEM\_KMII – Management IS/ICT

Semestrální práce

**Informační systém pro správu logů**Log Management – Elasticstack

**Student:** Bc. Václav Šulc  
**UČO:** 45552  
**E-mail:** [45552@mail.ambis.cz](mailto:45552@mail.ambis.cz)  
**Semestr:** LS 2021

Obsah

[1. Úvod 3](#_Toc64743288)

[1.1. Log Management 3](#_Toc64743289)

[2. Komponenty Log Management 4](#_Toc64743290)

[2.1. Elasticsearch 4](#_Toc64743291)

[2.2. Logstash 7](#_Toc64743292)

[2.3. Kibana 8](#_Toc64743293)

[2.4. Cerebro 13](#_Toc64743294)

[2.5. Redis fronta 13](#_Toc64743295)

[2.6. ReadonlyREST (ROR) 14](#_Toc64743296)

[2.7. Agenti zajišťující sběr dat 14](#_Toc64743297)

[3. Architektura 15](#_Toc64743298)

[3.1.1. Komunikace v Log Management clusteru 15](#_Toc64743299)

[3.1.2. Používaná architektura v podniku 16](#_Toc64743300)

[4. Závěr 18](#_Toc64743301)

1. Úvod

Tato práce se zabývá informačním systémem pro centralizovanou správu logů. Tento systém se označuje slovním spojením Log Management. Systém je postavený na produktech Elasticstack, který je vyvíjený společností [Elasticsearch B.V.](https://www.elastic.co/) Softwarové komponenty toho systému jsou tvořeny převážně open-source produkty a produkty, které jsou zdarma a mohou být používané i pro komerční použití bez nutnosti dokupování licencí (Elastic Stack subscriptions, 2021).

Log Management lze chápat jako aktivity a procesy používané pro generování, shromažďování, centralizaci, analýze, přenosu, ukládání, archivaci a likvidaci obrovských objemů počítačově generovaných logů (What Is Log Management?, 2019).

Log Management umožňuje podniku velmi rychle a efektivně detekovat chyby v operačních systémech a aplikacích. Systém poskytuje detailní přehled o činnostech uživatelů a zařízení. Systém také umožňuje splnit požadavky Zákona o kybernetické bezpečnosti a GDPR (Solnicki, 2021).

V podniku může být tento systém využíván ICT administrátory serverů a aplikací, bezpečnostními experty, analytiky nebo programátory.

* 1. Log Management

Log Management zajišťuje vyčítání, příjem, modifikaci, uložení, interpretaci a analýzu logů. Řešení postavené na Elasticstack se skládá z těchto komponent: *Elasticsearch, Logstash, Kibana, Beats a X-Pack.* Elasticstack je navíc obohacen o další programy jako je např. Cerebro nebo Redis fronta.



Jako databáze pro ukládání dat se používá tzv. Elasticsearch. Jde o distribuovanou databázi, kterou lze provozovat na vlastním HW nebo ve vlastním Cloudu. Nabízí se však i varianta využití Log Managementu jako služby, kterou poskytuje společnost Elasticsearch B.V.

Co se týče nabízení Log Managementu jako služeb, existují konkurenční služby jako [Loggly](https://www.loggly.com/), [Logz](https://logz.io/) či [Splunk](https://www.splunk.com/). V případě použití cloudových služeb je nutné brát v potaz, že přenos dat do těchto platforem je nutné řádně zabezpečit např. šifrovaným SSL/TLS spojením. S tím je spojena vyšší HW náročnost, jelikož šifrování a dešifrování stojí nějaký výkon, vyšší konfigurační pracnost. Do toho je také nutné započítat administrativu spojenou s certifikačními autoritami.

V případě, že se podnik rozhodne provozovat Log Management sám, je doporučeno využít automatizačních nástrojů na správu serverů a běžících služeb. Lze využít např. [Ansible](https://www.ansible.com/) nebo [Puppet](https://puppet.com/).

Na Log Management lze napojit alertovací nástroje, které podle zadaných pravidel filtrů čtou data a prohledávají logy v databázích a vyvolávají určité aletry ať už v podobě zaslání emailové notifikace, SNMP trapu nebo jiného typu alertu. Lze využít open-source nástroj [Elastalert](https://elastalert.readthedocs.io/en/latest/).

Log Management je nutnou komponentou pro budování tzv. [SIEM](https://www.elastic.co/siem) (Security Information and Event Management) řešení.

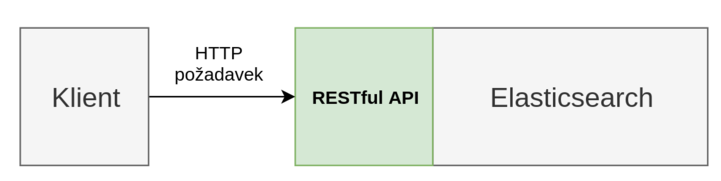
1. Komponenty Log Management
   1. Elasticsearch

Elasticsearch je distribuovaný databázový systém, který lze horizontálně škálovat. Výhodou je, že v případě nedostatku výkonu stačí přidat další server do tzv. Elasticsearch clusteru a tím se navýší prostředky (RAM, CPU, diskový prostor) pro celý systém. Systém řízení báze dat (SŘBD) v angličtině je znám jako *Distributed database management system (DDBMS).* DDBMS je tosoftware, který spravuje a distribuuje databáze tak, že je toto chování pro uživatele transparentní (Tan, 2009, s. 894).

Běžící instanci Elasticsearch se říká Elasticsearch node neboli Elasticsearch uzel. Pro slovo Elasticsearch se běžně používá zkratka ES. Elasticsearch clusterem se rozumí jedna nebo více běžících Elasticsearch instancí, které spolu komunikují a sdílejí data.

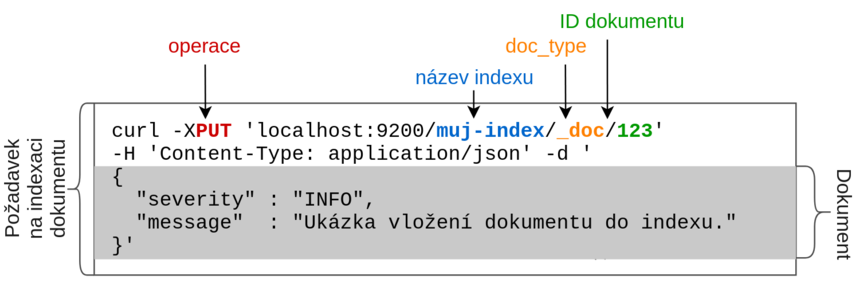
Elasticsearch lze také nazvat NoSQL databázi (Brasetvik, 2013), která využívá Java knihovny Apache Lucene (Celebrating 20 years of Apache Lucene, 2021); (Linwood, 2004, s. 255). Tento produkt vznikl přepsáním programu Compass. Elasticsearch poskytuje REST API, pomocí kterého lze databázi Elasticsearch spravovat (REST APIs, 2021). Elasticsearch lze volat pomocí jakéhokoliv programovacího jazyku (nejenom v Javě). Autorem tohoto programu je Shay Banon.

Zápis dat do databáze Elasticsearch se provádí právě přes zmíněné RESTful API (Vohra, 2015, s. 175). Zapisujícím klientem může být jakákoliv aplikace, které umí komunikovat pomocí HTTP protokolu.



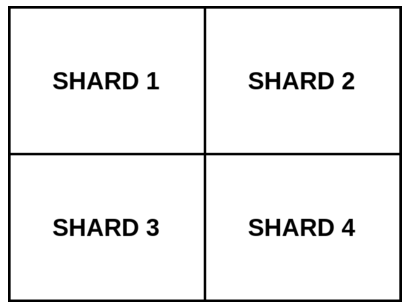
Obrázek 1: Ukázka zaslání požadavku na RESTful API

Obrázek níže detailněji popisuje požadavek:



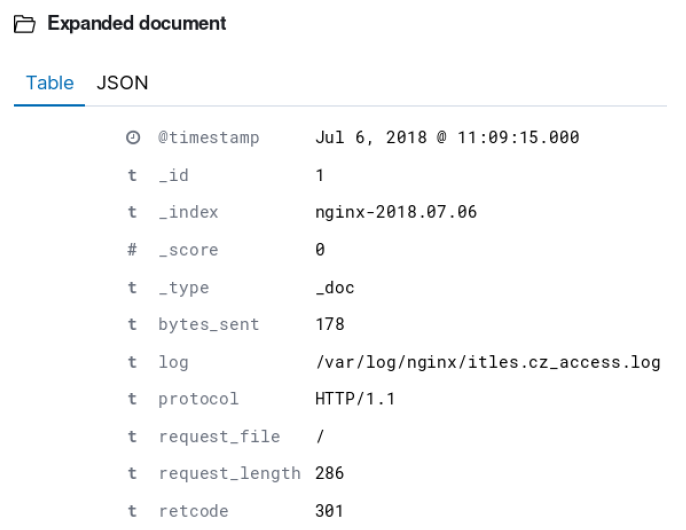
Obrázek 2: Detailní pohled na HTTP požadavek

Elasticsearch ukládá data do tzv. indexů. Index je logický jmenný prostor, který se dělá na menší části tzv. shardy, které lze pak distribuovat napříč server, které jsou připojeny do Elasticsearch clusteru (Sharma, 2016).



Obrázek 3: Rozdělení indexu na shardy

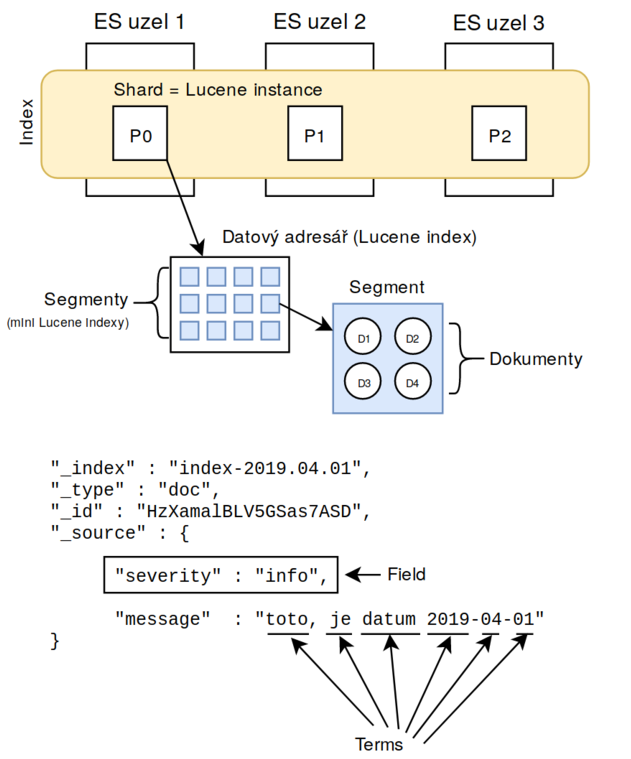
Shardy obsahují tzv. dokumenty. Tento dokument může obsahovat informace vyvolané operačním systémem nebo aplikací. Jde např. o IP adresu zařízení, závažnost zprávy, teplota, typ čidla, počet přenesených bytů, název uživatele apod.) Dokument je ve formátu JSON. JSON lze chápat jako formát pro výměnu dat (Smith, 2015, s. 37)



Obrázek 4: Pohled na Elasticsearch dokument v grafickém rozhraní Kibana



Obrázek 5: Pohled na Elasticsearch dokument v JSON formátu

Celkový pohled na Elasticsearch index, shard a dokument zobrazuje obrázek níže.

Obrázek 6: Grafické zobrazení pojmů Elasticsearch uzel, index, shard a dokument.

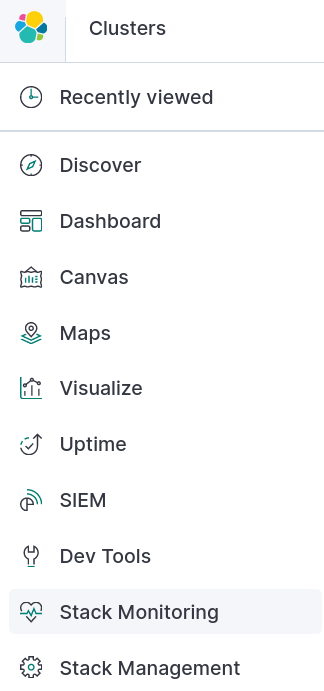
* 1. Logstash

Logstash je nástroj, který zajišťuje modifikaci, filtraci, přijímání, vyčítání a odesílání dat (Logstash, 2021). Data lze obohatit o určitá pole. Toto obohacení lze provést i podmíněně – např. pokud zpráva bude obsahovat určitá klíčová slova, lze ji kategorizovat jako určitý typ události. Některé události lze zahazovat a zabránit tak jejich uložení do databáze. Pomocí určitých filtrů lze události očišťovat (např. lowercase, odstranění přebytečných mezer apod.)

* 1. Kibana

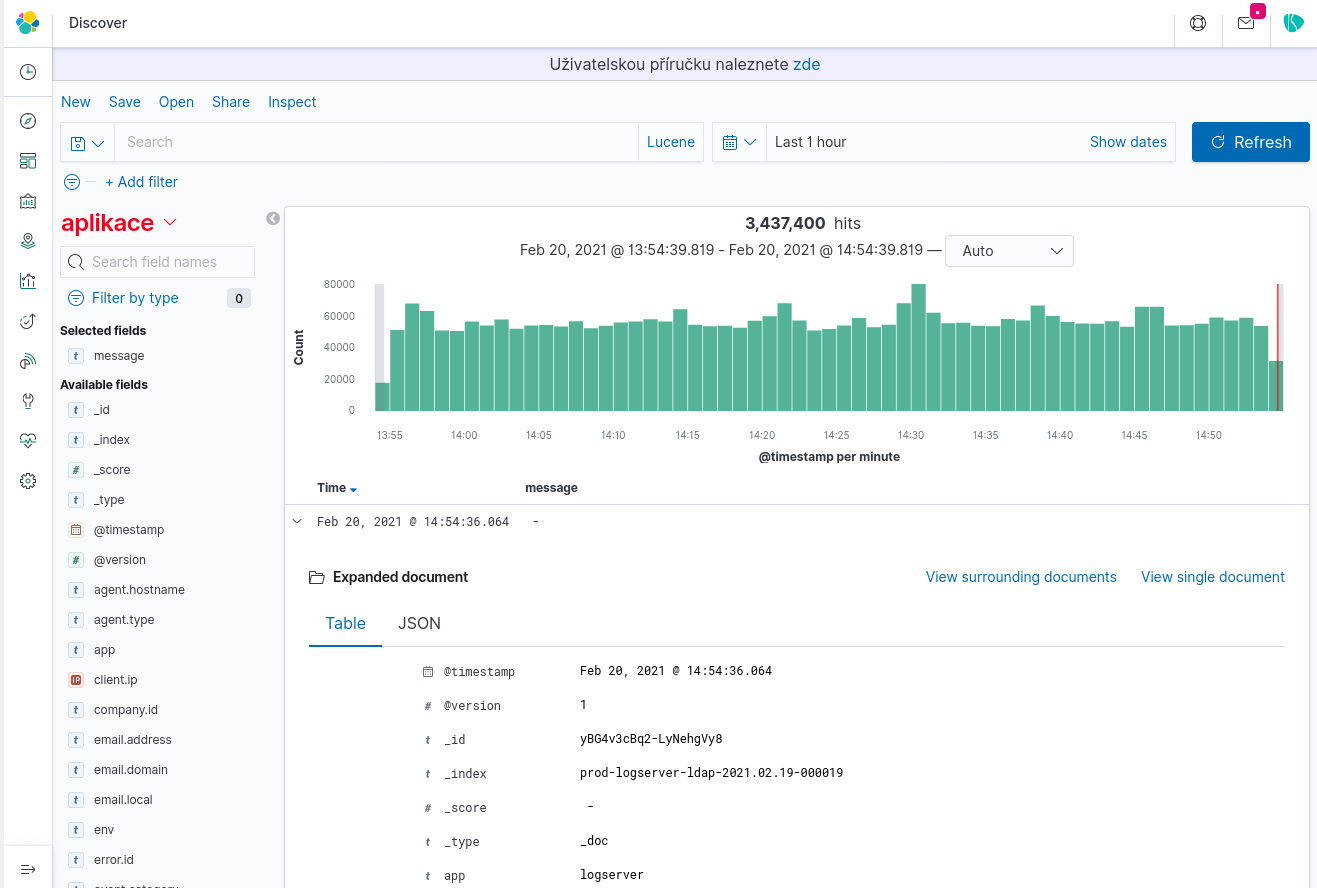
Je to grafické webové rozhraní, které slouží pro zobrazení a analýzu logů (Kibana, 2021). Webové rozhraní umožňuje prohlížení logů ve vybraných časových úsecích, umožňuje vytvářet vizualizace a dashboardy včetně reportů. Rozhraní také nabízí správu Elasticsearch clusteru a základní přehled o jeho zdraví a vytíženosti.

Níže je zobrazeno základní navigační menu:



Obrázek 7: Navigační menu webového rozhraní Kibana

V záložce Discover lze prohlížet logy (Discover, 2021), které jsou uloženy v NOSQL Elasticsearch databázi (clusteru).

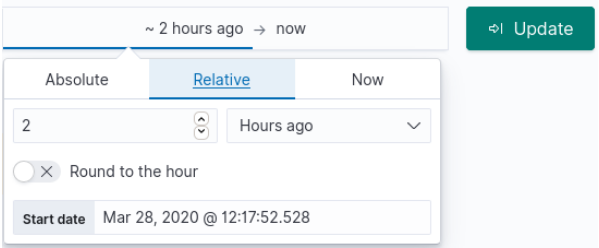


Obrázek 8: Ukázka záložky Discover ve webovém rozhraní Kibana

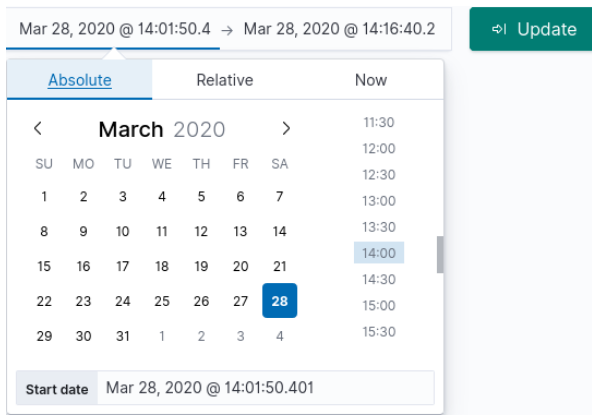


Obrázek 9: Ukázka záložky Discovery ve webovém rozhraní Kibana

Logy lze prohlížet ve stanoveném časovém intervalu. Lze využít relativního časového filtru, který se obvykle používá, když nevíme přesně, kdy k určité události došlo.

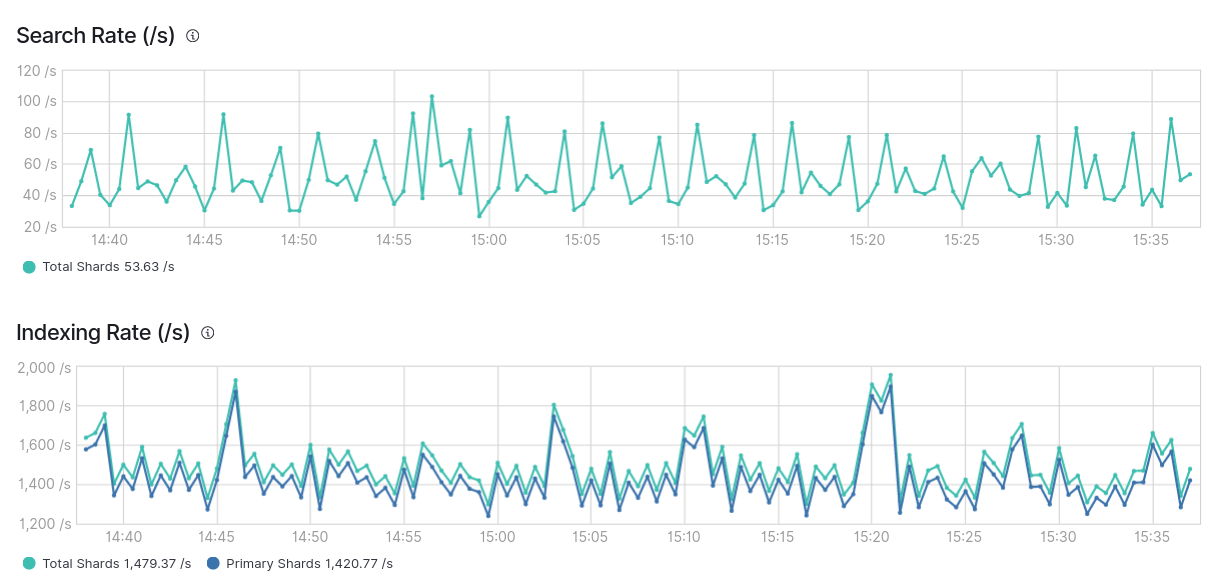


Obrázek 10: Ukázka relativního časového filtru ve webovém rozhraní Kibana

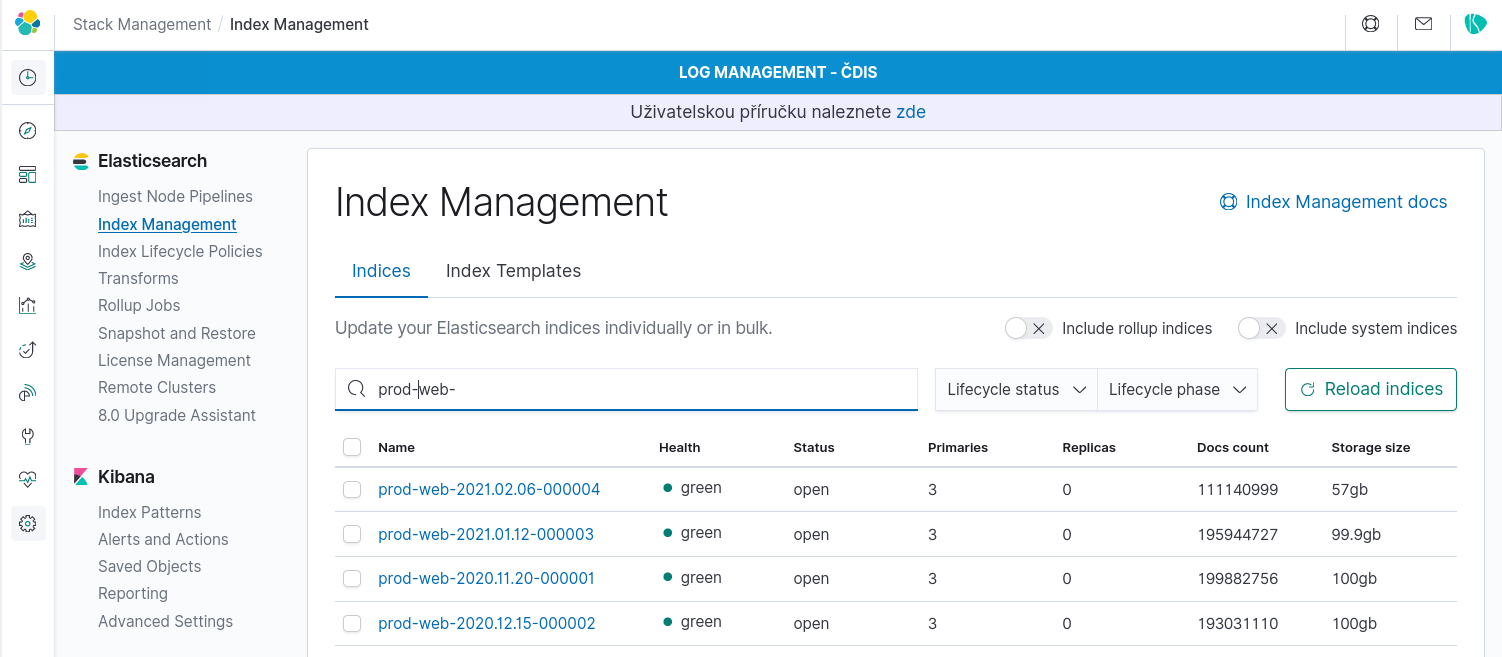


Obrázek 11: Ukázka absolutního časového filtru ve webovém rozhraní Kibana

Grafické rozhraní Kibana je určeno i pro ICT administrátory spravující Log Management cluster.

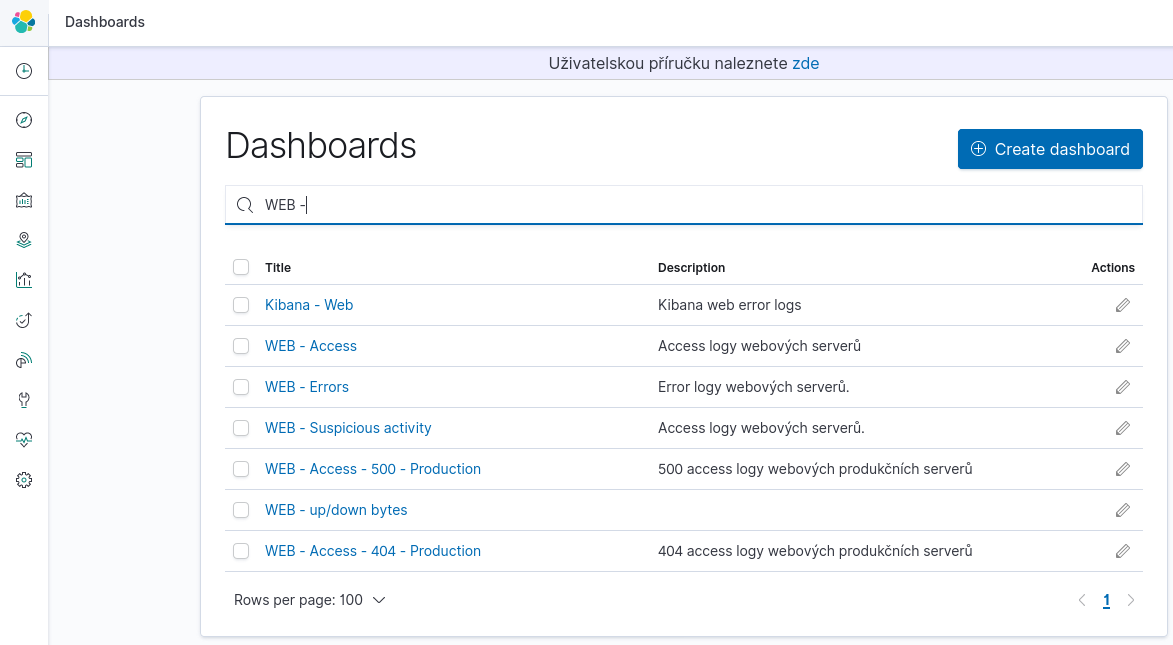


Obrázek 12: Diagnostické nástroje ve webovém rozhraní Kibana

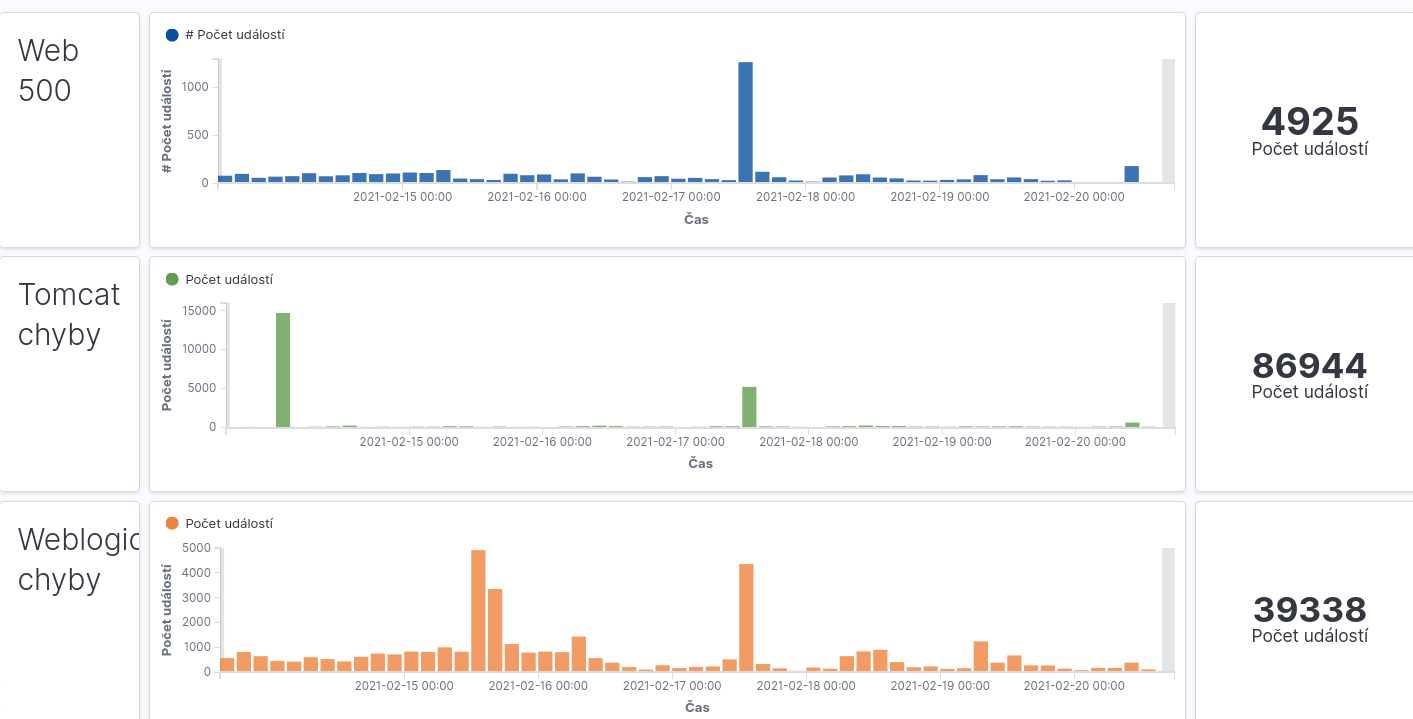


Obrázek 13: Správa indexů ve webovém rozhraní Kibana

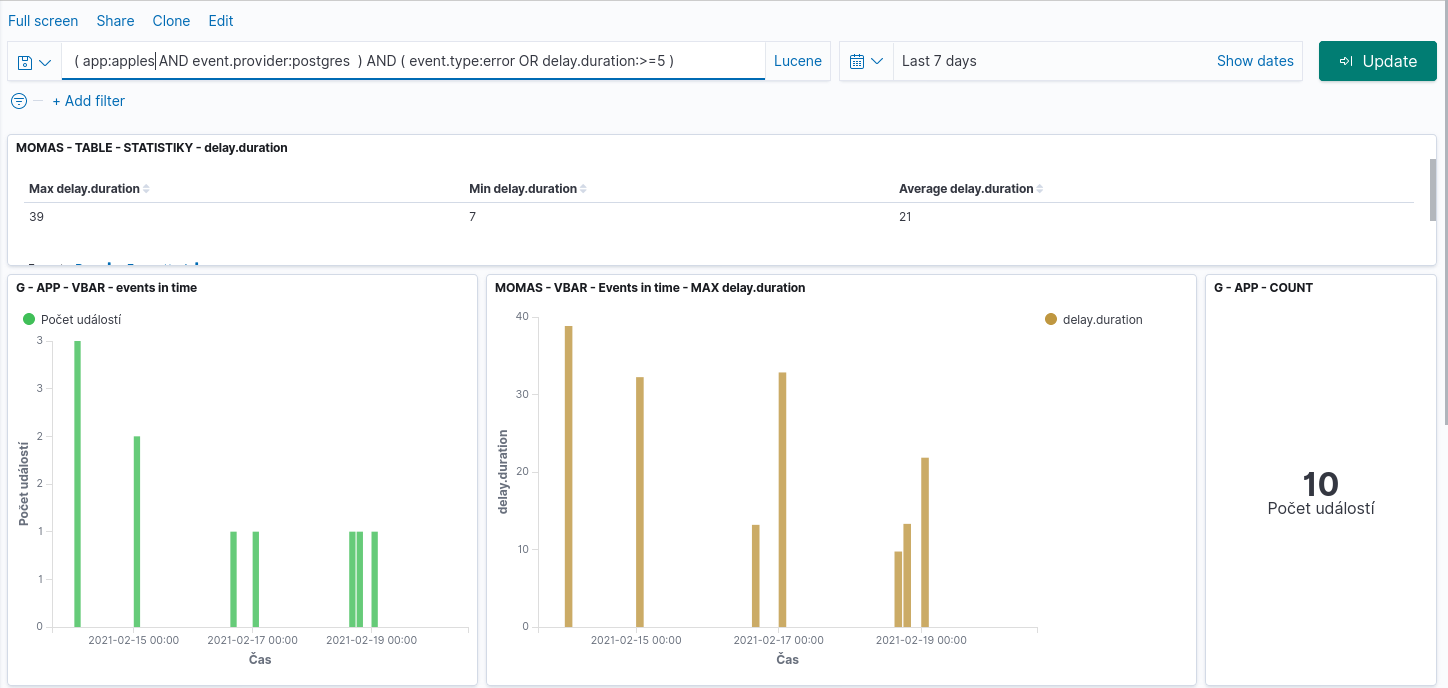
V Kibaně lze také tvořit dashbordy, které umožňují velmi rychlý pohled na určitou množinu informací. Může jít např. o dashboardy, který reprezentuje bezpečnostní události, nebo o dashbord, který sleduje chyby na aplikačních nebo webových serverech.



Obrázek 14: Ukázka dashboardů začínajících na WEB -



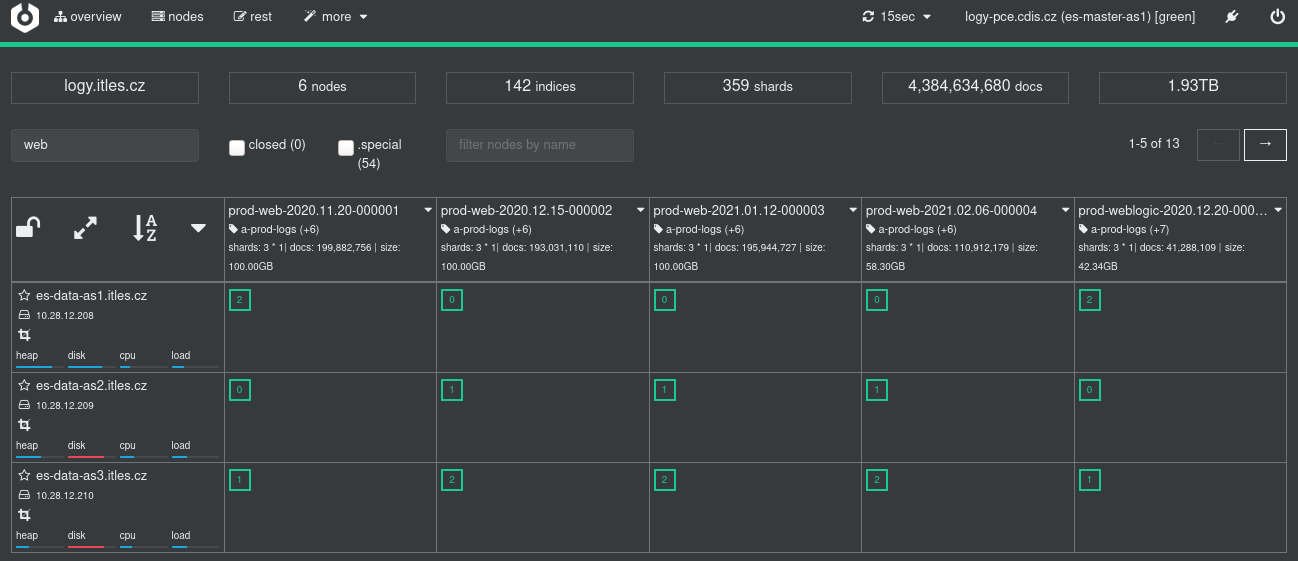
Obrázek 15: Dashboard reprezentující chyby ze 3 odlišných systémů



Obrázek 16: Dashboard sledující zpoždění vytvořené na aplikaci apples

* 1. Cerebro

Cerebro je webový nástroj, který umožňuje ICT administrátorovi spravovat data NoSQL databáze Elasticsearch na úrovni shardů (Cerebro, 2021). Rozhraní Cerebro dává přehledný grafický pohled na tzv. shardy, sleduje diskové vytížení serverů a jejich zátěž.



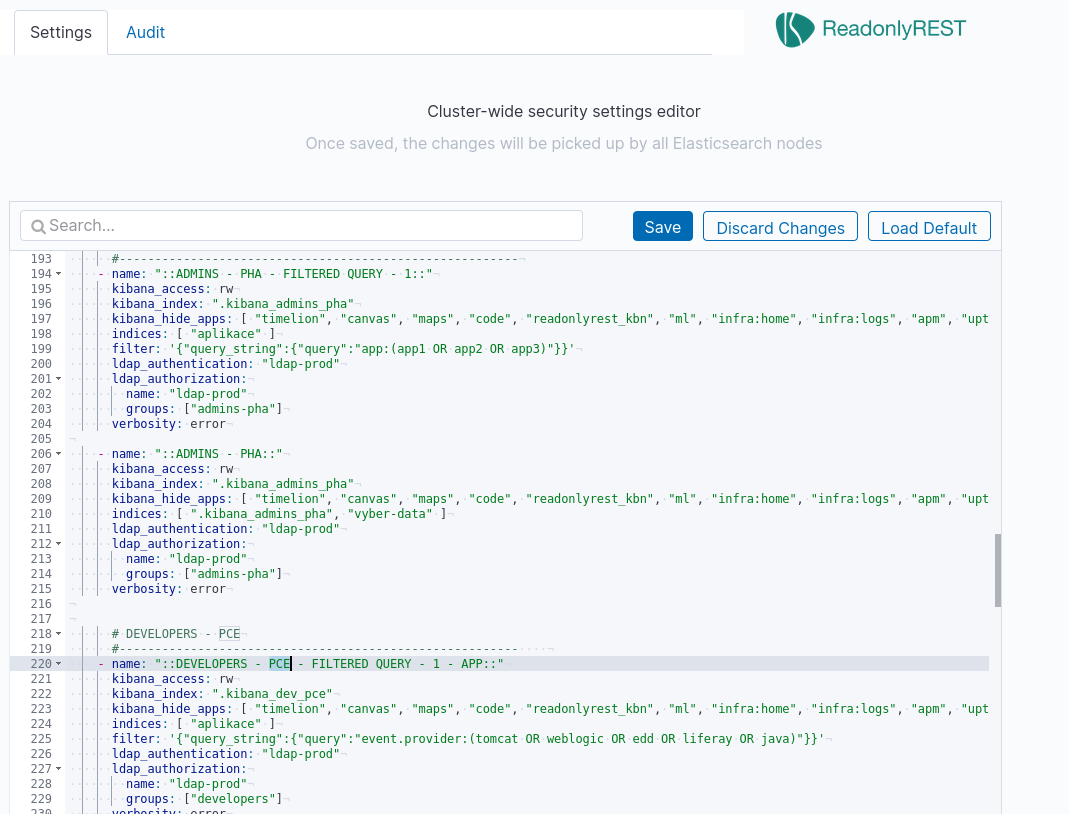
Obrázek 17: Cerebro - Webové rozhraní pro správu Elasticsearch databází

* 1. Redis fronta

Redis fronta zajišťuje bufferování událostí (Redis, 2021). To je užitečné v případě, kdy se systém upgraduje, když je nedostupný nebo přetížený. Redis fronta navíc umožňuje snadnější kategorizaci událostí, jelikož data do redis fronty lze uložit pod specifickým klíčem.

* 1. ReadonlyREST (ROR)

Jde o bezpečnostní plugin, který je možné nainstalovat do webového rozhraní Kibana a databáze Elasticsearch (Security for Elasticsearch and Kibana | ReadonlyREST, 2021). V základní variantě se poskytuje zdarma. Pokročilejší verze toho pluginu vyžadují placenou licenci (Pricing: Compare all the features, 2021).



Obrázek 18: Ukázka konfigurace bezpečnostního pluginu ReadonlyREST v grafickém rozhraní Kibana

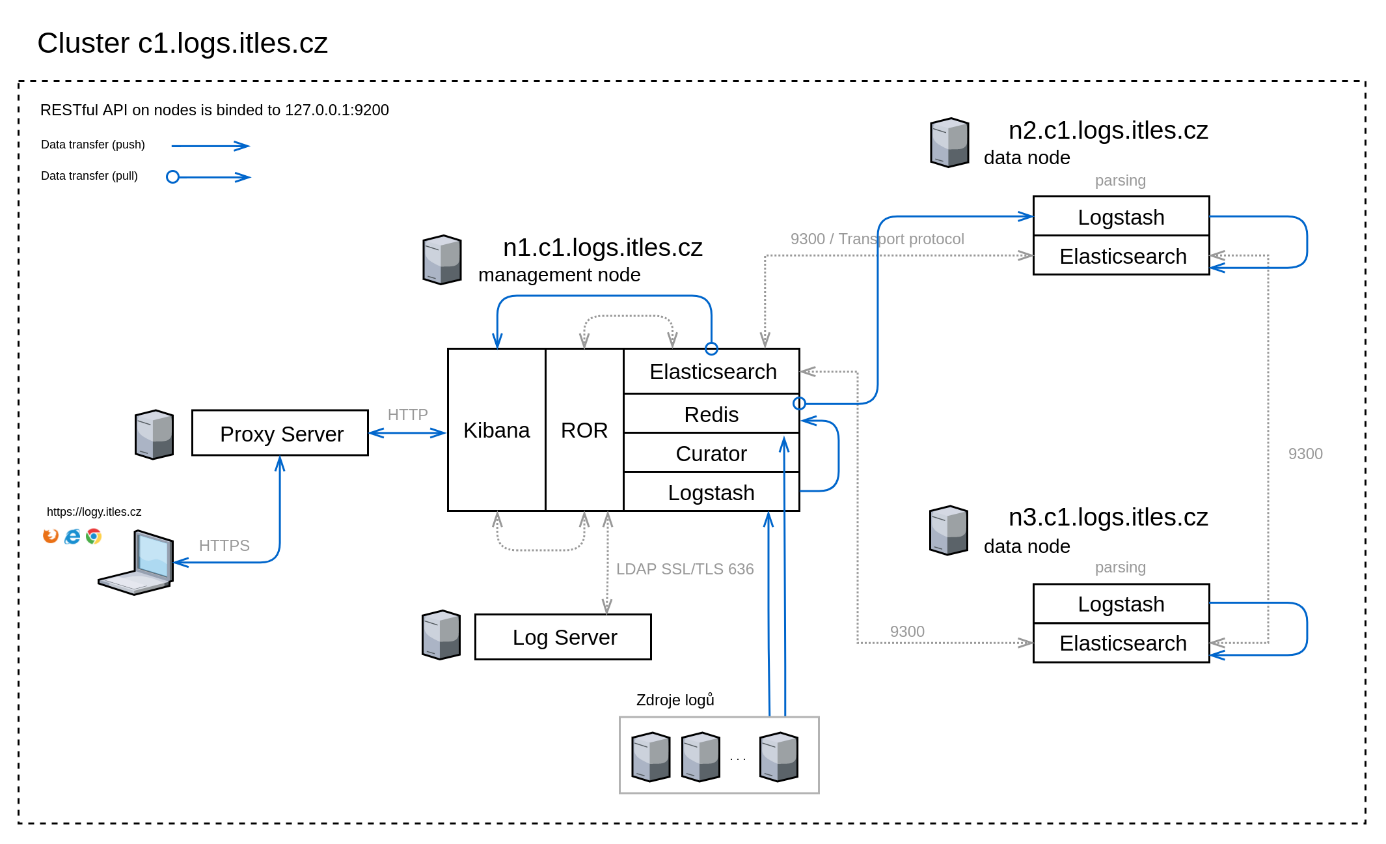
* 1. Agenti zajišťující sběr dat

V rámci Elasticstacku lze používat tzv. Beats agenty, které zajišťují sběr dat z operačních systémů (Lightweight data shippers, 2021). Jde např. o Filebeat (vyčítá data ze souborů na různých operačních systémech), Winlogbeat (vyčítá Windows Event události z operačních systémů Windows) a Auditbeat (vyčítá auditní informace z operačních systémů). Jako další možné nástroje zajišťující vyčítání dat ze souborů mohou být rsyslog (The rocket-fast Syslog Server, 2021) nebo syslog-ng (The rocket-fast Syslog Server, 2021).

Doporučuje se provést určité optimalizace na agentech, aby sběr dat příliš nezatěžoval operační systém, na kterém agent běží (Filebeat High CPU consumption on two of four hosts, 2019).

1. Architektura
   * 1. Komunikace v Log Management clusteru

Níže je graficky znázorněna základní architektura Log Management řešení. Tato architektura se snaží o minimalizaci HW prostředků. Na této architektuře budou popsány komponenty a jejich vzájemná komunikace.



Obrázek 19: Základní architektura Log Management řešení

Klient, který chce přistoupit na webové rozhraní Kibana ve skutečnosti přistupuje na reverzní proxy server, ten pak předává požadavek na samotné webové rozhraní Kibana. V okamžiku, kdy přijde požadavek na webové rozhraní Kibana – např. může jít o požadavek na zobrazení dashboardu s bezpečnostními událostmi, požadavek Kibana přepošle do bezpečnostního plugin ReadonyREST, který zajišťuje autentizaci a autorizaci uživatelů. ROR plugin je nakonfigurován tak, aby využíval služeb LDAP serveru (tj. Log Server ve schématu). ROR v podstatě slouží jako proxy mezi webovým rozhraním Kibana a Elasticsearch databází. Jakmile je uživatel úspěšně ověřen a má přístup k daným zdrojům, požadavek dál putuje do Elasticsearch clusteru. Část dat může být uložena na serveru n2.c1.cdis.cz a část na n3.c1.cdis.cz. Elasticsearch cluster si však sám zařídí a vyřídí veškeré náležitosti nutné pro získání dat z vybraných serverů. Tímto není uživatel ani správce zatěžován. Klient pak získá data, která se mu zobrazí ve webovém rozhraní Kibana.

Na serveru n1.c1.logs.itles.cz je také dostupné webové rozhraní Cerebro, které slouží pro administrátory pro správu Elasticsearch clusteru. Elasticsearch, který běží na serveru n1.c1.logs.itles.cz záměrně neukládá data a díky tomu není tento server zatěžován. To se dělá kvůli stabilitě clusteru.

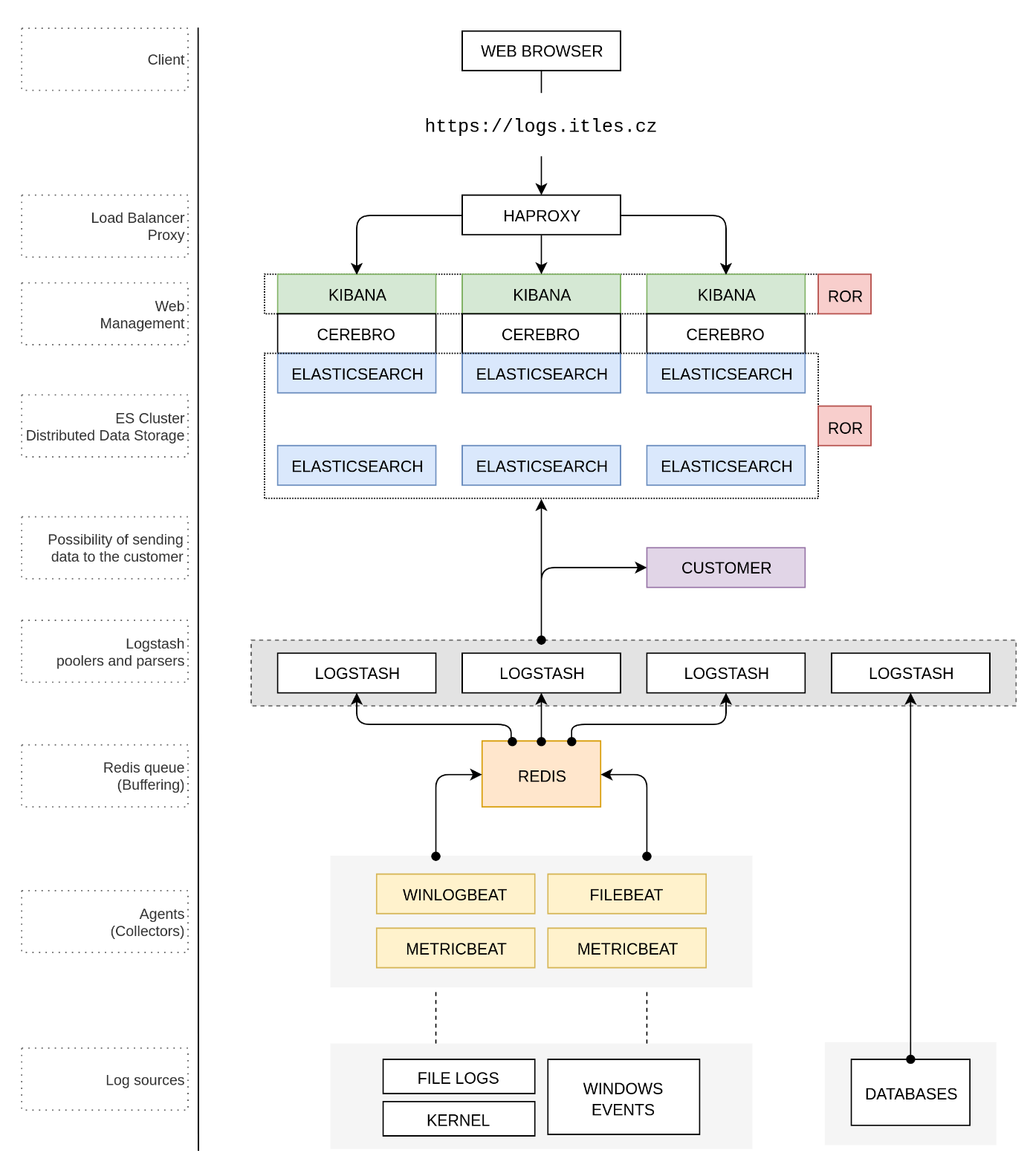
Na serveru n1.c1.logs.itles.cz je přítomna Redis fronta. Do této fronty se ukládají data z agentů a posléze se tyto data z fronty vyčítají programy Logstash. Logstash instance data mohou modifikovat, obohacovat o další informace (zprávy kategorizovat) nebo vybrané zprávy zahazovat. Samotné logy v podobě Elasticsearch dokumentů pak Logstash programy ukládají do Elasticsearch databází.

V případě, že by bylo potřeba navýšit HW prostředky např. z důvodu přeplňování Redis fronty, stačí přidat další server s instancí Logstash, který bude zajišťovat vyčítání zpráv z fronty. Kdyby byl problém s propustností při ukládání dat do Elasticsearch clusteru, stačí přidat další běžící instanci Elasticsearch do clusteru, které bude umožněno ukládat data.

Data napříč Elasticsearch uzly lze replikovat, a dokonce i zálohovat bez nutnosti externího zálohovacího SW.

* + 1. Používaná architektura v podniku

V podniku je použita níže uvedená architektura, která je proti základní architektuře více robustní a reflektuje požadavky na zvyšující se míru integrovaných logů z mnoha systémů. Architektura také zajišťuje vysokou dostupnost pomocí reverzní proxy, redundantních komponent a replikaci dat.



Obrázek 20: Aktuálně používaná architektura Log Management řešení v podniku.

1. Závěr

Informační systém Log Management napomáhá podnikům rychle reagovat na výskyty chyb v systémech a aplikacích. To firmám umožňuje řešit vzniklé problémy dříve, než dorazí stížnost ze strany zákazníka. Log Management lze také využít jako nástroj pro řešení kritických ICT událostí a pro snadnější plnění požadavků kybernetického zákona a GDPR.  
S ohledem na rostoucí míru digitalizace a využívání ICT prostředků, doporučuji podnikům se tomuto tématu blíže věnovat.

Před samotnou implementací tohoto systému doporučuji sepsat důvody, které podpoří návrh na realizaci toho systémů. Je nutné vydefinovat systémy, z kterých se budou data sbírány a jakým způsobem. Podnik by měl vědět jaké události sbírat a za jakým účelem. S tím je potřeba vydefinovat, které informace se budou z logu extrahovat. Do Log Management systému lze zapojit alertovací nástroj, který lze propojit s monitorovacími systémy pro rychlou detekci problémů a bezpečnostních incidentů. Doporučuje se také dodržovat standardizované názvy polí, atributů událostí, které pak umožňují korelovat události mezi více systémy (Elastic Common Schema (ECS) Reference, 2021). To lze velmi dobře využít pro korelování bezpečnostních událostí v pokročilých bezpečnostních systémech SIEM.

**Použité zdroje**

BRASETVIK, Alex, 2013. Elasticsearch as a NoSQL Database. *Free and Open Search: The Creators of Elasticsearch, ELK & Kibana | Elastic* [online]. Elasticsearch B.V. [cit. 2021-02-20]. Dostupné z: https://www.elastic.co/blog/found-elasticsearch-as-nosql

Celebrating 20 years of Apache Lucene, 2021. *Free and Open Search: The Creators of Elasticsearch, ELK & Kibana | Elastic* [online]. Elasticsearch B.V. [cit. 2021-02-20]. Dostupné z: https://www.elastic.co/celebrating-lucene

Cerebro, 2021. *GitHub* [online]. [cit. 2021-02-20]. Dostupné z: https://github.com/lmenezes/cerebro

Discover, 2021. *Free and Open Search: The Creators of Elasticsearch, ELK & Kibana | Elastic* [online]. Elasticsearch B.V. [cit. 2021-02-20]. Dostupné z: https://www.elastic.co/guide/en/kibana/current/discover.html

Elastic Stack subscriptions, 2021. *Elasticsearch B.V. All Rights Reserved* [online]. Elasticsearch B.V. [cit. 2021-02-20]. Dostupné z: https://www.elastic.co/subscriptions

Filebeat High CPU consumption on two of four hosts, 2019. *Discuss the Elastic Stack - Official ELK / Elastic Stack, Elasticsearch, Logstash, Kibana, Beats and More forums* [online]. [cit. 2021-02-20]. Dostupné z: https://discuss.elastic.co/t/filebeat-high-cpu-consumption-on-two-of-four-hosts/183399

Kibana, 2021. *Free and Open Search: The Creators of Elasticsearch, ELK & Kibana | Elastic* [online]. Elasticsearch B.V. [cit. 2021-02-20]. Dostupné z: https://www.elastic.co/kibana

Lightweight data shippers, 2021. *Free and Open Search: The Creators of Elasticsearch, ELK & Kibana | Elastic* [online]. Elasticsearch B.V. [cit. 2021-02-20]. Dostupné z: https://www.elastic.co/beats/

LINWOOD, Jeff a Dave MINTER, 2004. Integrating the Lucene Search Engine. LINWOOD, Jeff a Dave MINTER. *Building Portals with the Java Portlet API* [online]. Berkeley, CA: Apress, s. 255-280 [cit. 2021-02-20]. ISBN 978-1-59059-284-7. Dostupné z: doi:10.1007/978-1-4302-0754-2\_10

Logstash: Collect, Parse, Transform Logs, 2021. *Free and Open Search: The Creators of Elasticsearch, ELK & Kibana | Elastic* [online]. Elasticsearch B.V. [cit. 2021-02-20]. Dostupné z: https://www.elastic.co/logstash

Pricing: Compare all the features, 2021. *Security for Elasticsearch and Kibana | ReadonlyREST* [online]. Beshu Limited t/a ReadonlyREST Security [cit. 2021-02-20]. Dostupné z: https://readonlyrest.com/#pricing

Redis, 2021. *Redis Labs* [online]. [cit. 2021-02-20]. Dostupné z: https://redislabs.com/ebook/part-1-getting-started/chapter-1-getting-to-know-redis/

REST APIs, 2021. *Free and Open Search: The Creators of Elasticsearch, ELK & Kibana | Elastic* [online]. Elasticsearch B.V. [cit. 2021-02-20]. Dostupné z: https://www.elastic.co/guide/en/elasticsearch/reference/current/rest-apis.html

*Security for Elasticsearch and Kibana | ReadonlyREST* [online], 2021. Beshu Limited t/a ReadonlyREST Security [cit. 2021-02-20]. Dostupné z: https://readonlyrest.com/

SHARMA, Vishal, 2016. Getting Started with Elasticsearch. SHARMA, Vishal. *Beginning Elastic Stack* [online]. Berkeley, CA: Apress, s. 17-28 [cit. 2021-02-20]. ISBN 978-1-4842-1693-4. Dostupné z: doi:10.1007/978-1-4842-1694-1\_2

SMITH, Ben, 2015. Introducing JSON. SMITH, Ben. *Beginning JSON* [online]. Berkeley, CA: Apress, s. 37-47 [cit. 2021-02-20]. ISBN 978-1-4842-0203-6. Dostupné z: doi:10.1007/978-1-4842-0202-9\_4

SOLNICKI, Maciej, 2021. ČSN/ISO 27001:2013 a LOGmanager. *Log Manager* [online]. [cit. 2021-02-20]. Dostupné z: https://www.logmanager.cz/blog-post/csn-iso-270012013-a-and-logmanager/

TAN, Kian-Lee, 2009. Distributed Database Systems. LIU, LING a M. TAMER ÖZSU, ed. *Encyclopedia of Database Systems* [online]. Boston, MA: Springer US, s. 894-896 [cit. 2021-02-20]. ISBN 978-0-387-35544-3. Dostupné z: doi:10.1007/978-0-387-39940-9\_701

The rocket-fast Syslog Server, 2021. *Rsyslog* [online]. Adiscon GmbH [cit. 2021-02-20]. Dostupné z: https://www.rsyslog.com/

VOHRA, Deepak, 2015. Using Elasticsearch. VOHRA, Deepak. *Pro Couchbase Development* [online]. Berkeley, CA: Apress, s. 175-196 [cit. 2021-02-20]. ISBN 978-1-4842-1435-0. Dostupné z: doi:10.1007/978-1-4842-1434-3\_7

What Is Log Management?, 2019. *Graylog* [online]. [cit. 2021-02-20]. Dostupné z: https://www.graylog.org/post/what-is-log-management-a-complete-logging-guide