

Tytuł szkolenia: Elasticsearch Logstash Kibana

Kod szkolenia: ELK

Wprowadzenie

Czas trwania: 1 do 4 dni, w zależności od aktualnego doświadczenia uczestników z Elasticsearch. Na przerobienie całego konspektu szkolenia zaleca się 4 dni, ale wybrane tematy mogą być również przeprowadzone w ciągu 1 dnia.

Szkolenie może zostać przeprowadzone zarówno w języku polskim jak i angielskim oraz może być dostosowane do indywidualnych potrzeb i wymagań grupy.

Adresaci szkolenia

Cel szkolenia

- Umiejętność konfiguracji klastra Elasticsearch o parametrach odpowiednich dla potrzeb projektowanego systemu
- Umiejętność praktycznego użycia Elasticsearch i integracji z istniejącym oprogramowaniem (unikanie tzw. golden hammer),
- Umiejętność projektowania optymalnego modelu danych
- Efektywne mechanizmy ingest i query
- Umiejętność przetwarzania i transformowania danych
- Umiejętność pracy z NoSQL
- Umiejętność pracy w rozproszonych systemach
- Umiejętność integracji SQL i NoSQL oraz optymalizacja wydajności
- Testowanie wydajności rozproszonego systemu i odnajdywania wąskich gardeł,
- Umiejętność wyboru gotowych na rynku narzędzi i frameworków celem optymalizacji pracy developera – użycie technologii które do tej pory nie były brane pod uwagę a z sukcesem mogą zostać wdrożone

Cena kursu różni się w zależności od ilości dni szkolenia. **Skontaktuj się z nami, aby otrzymać dedykowaną ofertę .**

Czas i forma szkolenia

- 14 godzin (2 dni x 7 godzin), w tym wykłady i warsztaty praktyczne.

Plan szkolenia

1. Wprowadzenie

- Dlaczego świat ekscytuje się BigData i co to jest? – *wprowadzenie w świat technologii BigData*
- Własny DC czy Chmura – przegląd plusów i minusów w kontekście Amazon Web Services (AWS)
- Przegląd dostępnych technologii OpenSource BigData (tj.: Hadoop, Hbase, Hive, Pig, Kafka, RabbitMQ, Elasticsearch itp.) – *jakich produktów należy używać i kiedy, jakie są różnice między nimi, jakie są wady i zalety poszczególnych z nich*
- Open Source vs Closed Source – *dlaczego warto wybierać technologie OpenSource*
- NoSQL vs SQL – *wprowadzenie do świata NoSQL, wady i zalety, transakcyjność, systemy rozproszone*

2. Elasticsearch – podstawy

- Indeksy Lucene – *co to jest i dlaczego mówimy o Lucene w kontekście Elasticsearch*
- Architektura produktu Elasticsearch
- Środowisko pracy – *Window vs Unix, różne wersje Javy*
- Node, shard, replica, mappings, routing, alokacja, templates, dynamic mapping – *podstawowe elementy Elasticsearch , których zrozumienie jest kluczowe*
- Jak Elasticsearch przechowuje dane – *objects vs nested objects, refresh, cache, source, store, multi fields, wersjonowanie,*

definiowanie kluczy, wewnętrzne struktury danych + mechanizmy cache'owania, inverted index, cache

3. Elasticsearch – standalone instance

- Uruchomienie standalone instance
- Index, update, upsert – tworzenie i aktualizacja pojedynczych dokumentów oraz poprzez bulk
- Search – przegląd Query DSL
- Query, filter, aggregation(facets) – podstawy zapytań NoSQL czyli jak efektywnie pytać o dane
- Rating, shard i formuły liczenia scoringu – $TF*IDF$, $BM25$ czy może własny?

4. Elasticsearch – budowanie klastra

- Skalowalność – Vertical scaling vs Horizontal scaling
- Mechanizm Zen Discovery
- Konfiguracja nodów w klastrze (ingest/data/master/ml), dobór ilości masterów w klastrze
- Fault tolerance – odporność na awarie
- Optymalizacja klastra – parametry konfiguracyjne, custom allocation, tags, rack_id, zone itp.
- Split-brain effect, load balancing, transport client – czyli podstawowe elementy architektury o których należy pamiętać
- Study Case – ile klastrów chcemy mieć i jak zasilać je danymi

5. Logstash + Beats

- Uruchomienie i konfiguracja produktu
- Transformacja danych
- Ingest do Elasticsearch

6. Kibana

- Architektura UI w kontekście Elasticsearch – jak najlepiej zamodelować UI
- Kibana (discovery, metryki, index patterns, devtools, dashboardy, canvas, APM server, machine learning, X-Pack)

7. Elasticsearch – mechanizmy zaawansowane

- Query vs Ingestion – zaawansowane mechanizmy populacji danych poprzez systemy kolejkowe t.j Kafka, RabbitMQ
- NoSQL Query – budowanie zaawansowanych zapytań, dlaczego każde kolejne jest szybsze?
- Ingestion – jak efektywnie indeksować TB danych w Elasticsearch
- Zaawansowane konfiguracja klastra Elasticsearch – przegląd i wyjaśnienie parametrów konfiguracyjnych
- Indexer, data storage czy reporting – jak dostosować klastr do naszych potrzeb
- Tokenizer, Analyser, Filter, Char filter – jak budować i testować zaawansowane analizatory i tokenizery dokumentów
- Machine Learning – detekcja anomalii
- Percolate – reverse search, system alertów
- Suggesters, spell corrections, highlighting
- Search – profile, explain, custom scoring, effective pagination
- Monitorowanie klastra – zapoznanie z Kibaną oraz alternatywnymi podejściami,
- Testowanie klastra pod kątem wydajności – optymalizacja wydajności i profilowanie
- Pluginy – jakie są dostępne i jak tworzyć własne(plusy i minusy)
- X-Pack – opis możliwości
- Security

8. Budowanie UI

- Architektura UI w kontekście Elasticsearch – jak najlepiej zamodelować UI
- Użycie gotowych UI – wady i zalety
- Kibana – przegląd możliwości
- Bezpieczne wyciąganie danych z klastra – czyli jak nie dopuścić do stanu „red”

9. Codziennosc z Elasticsearch

- Integracja Elasticsearch z istniejącymi systemami
- Typowe problemy utrzymania klastra, synchronizacji, skalowalności – dyskusja, podzielenie się doświadczeniami z wymagających projektów BigData

- Backupy
- Kierunki rozwoju Elasticsearch i kompatybilność wsteczna – a właściwie jej brak, czyli jak sobie z tym radzić (dla przykładu: TransportClient vs RESTclient, plugins vs no-plugins itp.)

10. Rozszerzony program warsztatów