kod szkolenia: ES_INT / PL AA 2d

Elasticsearch: Intermediate

Przeznaczenie szkolenia

Szkolenie skierowane do administratorów, programistów, testerów oraz wszystkich osób związanych z Elasticsearch chcących nauczyć się lub podnieść kompetencje. Szkolenie wymaga od uczestników znajomości produktu Elasticsearch w stopniu przynajmniej podstawowym.

Korzyści wynikające z ukończenia szkolenia

Jest to kolejne szkolenie z cyklu szkoleń o Elasticsearch, będące następnym etapem w budowaniu kompetencji z zakresu zarządzania, konfigurowania oraz używania produktu Elasticsearch. Szkolenie porusza już bardziej zaawansowane aspekty produktu pozwalające na swobodne wdrażanie Elasticsearch jako elementu infrastruktury BigData w środowiskach produkcyjnych. Po zakończeniu szkolenia uczestnicy są w stanie samodzielnie konfigurować i optymalizować produkcyjne klastry Elasticsearch przetwarzające setki albo i tysiące TB danych.

Oczekiwane przygotowanie słuchaczy

- Znajomość zagadnień wchodzących w skład szkolenia Elasticsearch Essentials
- Podstawowa wiedza o serwisach RESTful
- Podstawy bash, w tym użycia curl (ewentualnie innego oprogramowania do wysyłania requestów HTTP jak Postman, Fiddler itp.)
- Java w stopniu podstawowym

Język szkolenia

Szkolenie: polski

Czas trwania

2 dni / 14 godzin

Agenda szkolenia

1. Wprowadzenie do Elasticsearch
   - Uruchomienie lokalnego klastra Elasticsearch,
   - Zasilanie klastra przykładowymi danymi,
   - Przegląd najczęściej używanych elementów query DSL,
2. Elasticsearch – agregacje (facets)
   - Przegląd najczęściej używanych typów agregacji
   - Ryzyko używania agregat,
   - Optymalizacje agregat,
3. Mapping zaawansowany
   - Przegląd możliwości mappingu – optymalizacja mappingu oraz jakie są korzyści z dobrze zdefiniowanego mappingu,
   - Jak przechowywać dane – analiza przypadku z wykorzystaniem nested objects, multi fields; wpływ na performance i wskazanie potencjalnych problemów.
   - Wpływ mappingu na wielkość danych,
   - Odpowiedniego wykorzystywania wewnętrznych struktur danych Elasticsearch celem budowania efektywnych zapytań (cache, inverted index, norms, source, store, doc_values, fielddata, stronicowanie, scan, scroll, scoring calculation itp.)
4. Elasticsearch cluster
   - Konfiguracja nodów w klastrze (ingest/data/master/ml) – dobór ilości masterów w klastrze, ilości nodów w klastrze, ilości indeksów oraz właściwy dobór ilości oraz wielkości shardów dla indeksu,
   - Parametry konfiguracyjne,
   - Mechanizm Zen Discovery
   - Skalowanie klastra Elasticsearch,
   - Cykle życia oraz stany indeksów,
   - Fault tolerance – odporność na awarie
   - Optymalizacja klastra – parametry konfiguracyjne, custom allocation, tags, rack_id, zone itp.
   - Split-brain effect, load balancing, transport client,
   - Study Case – ile klastrów i jak zasilać je danymi
   - Zaawansowane konfiguracja klastra Elasticsearch – przegląd i wyjaśnienie parametrów konfiguracyjnych
   - Cross-cluster search

5. Backupy
   - Jak wykonywać backup i restore klastra Elasticsearch,
6. Elasticsearch w praktyce
   - Integracja ElasticSearch z istniejącymi systemami
   - Massive ingestion,
   - Rozwiązywania typowych problemów dnia codziennego w pracy z Elasticsearch
   - Zarządzania klastrem poprzez RESTful API (routing, allocation)
7. Elasticsearch – analizatory danych,
   - Wyjaśnienie znaczenia index_analyzer oraz search_analyzer,
   - Tokenizer, Analyser, Filter, Char filter – jak budować i testować zaawansowane analizatory i tokenizery dokumentów
   - Budowanie dedykowanych analizatorów tekstu,
   - Przegląd dostępnych tokenizerów i filtrów
   - Budowanie zaawansowanych analizatorów tekstu z użyciem skryptów